DTIC FILE COPY

1

AGARD-LS-171

AGARD-LS-171

AD-A225 965

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

DTIC
ELECTE
AUG 3 0 1990
S D

AGARD LECTURE SERIES No. 17

Benefits of Computer Assisted Translation to Information Managers and End-Users

(L'Intérêt de la Traduction Assistée par Ordinateur
pour les Responsables de Centres d'Information
et pour les Utilisateurs Finaux)

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY
ON BACK COVER

90 08 29 055

AGARD-LS-171

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARD Lecture Series No.171

Benefits of Computer Assisted Translation to Information Managers and End-Users

(L'Intérêt de la Traduction Assistée par Ordinateur pour les Responsables
de Centres d'Information et pour les Utilisateurs Finaux)

Accesion For
NTIS CRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution /

Availability Codes

Dist	Avail and / or Special
A-1	

QUALITY INSPECTED 4

This material in this publication was assembled to support a Lecture Series under the sponsorship of the Technical Information Panel of AGARD and the Consultant and Exchange Programme of AGARD presented on 14th—15th June 1990 in Washington D.C., United States, 25th—26th June 1990 in Brussels, Belgium, and 28th—29th June 1990 in London, United Kingdom.

The Mission of AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;
- Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Exchange of scientific and technical information;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.

Published June 1990

Copyright © AGARD 1990
All Rights Reserved

ISBN 92-835-0568-9

Printed by Specialised Printing Services Limited
40 Chigwell Lane, Loughton, Essex IG10 3TZ

Abstract

The aim of this Lecture Series is to show how computer assisted translation (CAT) can be of benefit not only to information managers but also to end-users. Existing systems will be described as well as the nature of the texts to be processed, the technical and human problems related to the use of such systems and the needs of end-users (quality level of translations, information acquisition in the mother tongue. . .). Examples of on-going applications and systems under development will also be presented. These examples will highlight the benefits documentation centres will derive from CAT and suggest solutions of interest to the end-user.

This Lecture Series, sponsored by the Technical Information Panel of AGARD, has been implemented by the Consultant and Exchange Programme.

See p V

Abrégé

Ce Cycle de Conférences a pour but de montrer l'intérêt que peut apporter la Traduction Assistée par Ordinateur (TAO) non seulement pour le responsable d'un Centre d'Information, mais également pour l'utilisateur final. Après avoir défini les systèmes existants, la nature des textes à traiter, les problemes techniques et humains liés à l'utilisation des systèmes et les besoins des utilisateurs finaux (qualité des traductions, connaissance de l'information dans la langue maternelle...), des exemples d'application en cours ou en développement seront présentés. Ces diverses applications permettront de dégager l'intérêt que pourront en tirer les Centres de Documentation et de proposer des solutions au bénéfice de l'utilisateur final.

Ce Cycle de Conférences est présenté dans le cadre du Programme des Consultants et des Echanges, sous l'égide du Panel de l'Information Technique de l'AGARD.

List of Authors/Speakers

Lecture Series Director: Mr O.Lavroff
Aérospatiale
12, rue Pasteur
92152 Suresnes
France

AUTHORS/SPEAKERS

Mr D.Bostad
Technical Translation Directorate
Foreign Technology Div.
Wright-Patterson AFB
Ohio 45433
United States

Mrs J.Gordon
Management of Central Dept.
Official Languages and Translations
Secretary of State
Terrasse de la Chaudière
Ottawa K1A OM5, Ontario
Canada

Mr I.Pigott
Commission des Communautés Européennes
Bat. Jeran Monnet B4-24A
L-2920 Luxemburg

Mr D.Pinna
Responsable Service Traductions
Direction Après Vente
Division Aviation
Aérospatiale
31060 Toulouse Cedex 3
France

Mr T.Schneider
Siemens AG AP 323
Carl Wely Straße 22
D-8000 Munich 83
Federal Republic of Germany

Mr A.Yanez
CEDOCAR
26, Blvd Victor
00460 Paris Armèes
France

Partial Contents:

	Page
Abstract/Abrégé	iii
List of Authors/Speakers	iv
	Reference
Typology of Existing Systems by Ian M.Pigott	1
L'Environnement Technique de la Traduction Assistée par Ordinateur par A.Yanez	2
Essai de Typologie des Textes Source, dans le Cadre de la Traduction Assistée par Ordinateur par Mrs J.Gordon	3
Problems Encountered with the Use of Computer Assisted Translation in a Technical Publications Production Environment by D.J.Pinna	4
End Users' Needs by T.Schneider	5
Aspects of Machine Translation in the United States Air Force by Dale A.Bostad	6
The Processing of Technical Literature at Aérospatiale (Processing Time and Cost) by D.J.Pinna	7*
State of the Art in West Germany by T.Schneider	8
Synthèse des Solutions Proposées aux Utilisateurs par A.Yanez	9
Comparaison entre les Traductions Humaines et les Traductions Automatiques (Qualités, Coûts et Délais) Comparison between Human and Automatic Translations (Quality, Costs and Processing Time) and Appendix 1 par O.Lavroff	10
Bibliography	B

Keywords: NATO formatted (KP)

---

* Not available at time of printing.

TYPOLOGY OF EXISTING SYSTEMS

Ian M. Pigott
Head of Systran Project
Commission of the European Communities
Jean Monnet Building B4/24A
L-2920 Luxembourg

Abstract

Various attempts have been made at defining a typology of MT systems, some based on generations of software and hardware developments, others on the nature of the translation process (e.g. direct, transfer, modular). Today, however, a classification based more generally on performance and user access would appear to be more appropriate. The paper will thus distinguish between large software packages installed on mainframe computers for access by telecommunications and smaller PC packages functioning on MS-DOS equipment. Attention will also be given to systems capable of dealing with limited vocabulary and syntax as well as to developments in Japan which are beginning to set new trends in MT technology. Finally, information will be presented on how systems are now being used in practice and how use is likely to evolve over the next decade.

Introduction

The typology of machine translation systems has been discussed and rediscussed over the past ten to fifteen years. Initially, suppliers and research centres tended to equate the maturity of their developments in terms of "software generations" in much the same way as computer suppliers. Distinctions based on generations became less and less meaningful as time went by, particularly as some approaches labelled second, third or even fourth generation proved less reliable in practice than earlier developments which had continued to mature.

John Hutchins in "Machine Translation - past, present, future" bases his typology on the nature of the translation process itself. He thus distinguishes between direct (bilingual), interlingual, transfer and semantics-based systems. The problem here is that practically all major developments have tended to progress along similar lines. Systems which originally took a direct or bilingual approach have since evolved into interlingual or even transfer systems while "semantics-based" systems have begun to give additional attention to many of the syntactic criteria adopted in earlier developments.

It is for the above reasons that in presenting my own ideas on MT typology, I shall give more emphasis to performance, improvability and user-friendliness than to distinctions in the linguistic make-up of systems. Users are after all more interested in how well a system can do the job than in how the job is actually done.

Existing systems

Most of the systems in current use originated in the United States in the sixties and seventies. They fall into two basic categories: the larger, more complex systems such as Logos, Spanam and Systran which are normally installed on centralized mainframe computers and can be accessed by telecommunications; and less sophisticated products such as Smart, Globalink, Linguistic Products and Weidner which run on personal computers or workstations at the user site. This second category should however not be underestimated since in the language software industry, as in other areas, there is a general tendency for desktop applications to evolve rapidly on the basis of user requirements.

Recent newcomers to the user market include Metal and Tovna. Metal was originally developed by the University of Austin in Texas and is currently supported by Siemens, Munich. The system now runs on Unix and extensions from the original German-English are being made to cover Spanish, French and Dutch. Tovna, another Unix-based system, is being developed in Jerusalem and has already been installed at several user sites for English-French.

Finally, over the past couple of years a number of Japanese systems have reached the marketplace, mainly for Japanese-English and/or English-Japanese. However, Fujitsu's Atlas system is already being extended to European language combinations and other Japanese manufacturers are likely to follow this trend. Given the enormous investments now being made by all the large Japanese companies in machine translation and related technologies, products from Japan are likely to start penetrating the European and US markets within the next couple of years.

Quality of output depends very much on the language pairs involved, the type of document and, of course, the coverage of technical terminology. It often happens that a given product will provide a reasonable level of quality for one language pair and far less satisfactory results for another. As a general rule, developments involving the Latin languages (French, Spanish, Italian and Portuguese) and English tend to produce rather better results, for a given amount of investment, than systems involving Germanic or Asian languages.

But language coverage in MT systems is now generally very good. Most operational systems cover French and English in both directions and most also have German and Spanish as either a source or target language. English is undoubtedly the most highly developed source language with a wide range of targets such as Spanish, Dutch, Portuguese, Danish, Swedish, Japanese and Arabic. Russian, an old favourite in the 1960s, is regaining attention along with Chinese and Korean which have joined the club more recently.

User requirements

User requirements fall basically into two categories: information assimilation and information dissemination, although of course there are grey areas between these two.

- Information assimilation can be described as the gathering of information from internal or external sources for general use by an individual in keeping abreast with evolving policies, markets or technical advances;

- Information dissemination covers the whole process of communicating or publishing documents for (often unidentified) third parties.

Examples of documents used or translated for purposes of information assimilation are press reports, documentary data bases, technical reports from consultants or from industry in general. The reader's main aim is to understand the message of the documents in question and he will thus usually accept comparatively lower standards of translation. Very often, in this context, speed and low cost are of primary importance.

In regard to assimilation, the United States Air Force have used Systran since 1970 to translate first from Russian and later from French and German into English. The documents cover a wide range of technical sectors and user satisfaction is said to be high. In Europe, the Nuclear Research Centre in Karlsruhe, West Germany, has a similar application involving the translation of French-language research papers into English. At the European Commission too, use of raw machine translation for information purposes has been steadily increasing over the past couple of years, particularly in cases where users are unable to obtain human translations within the time available.

As for information dissemination, documents currently being submitted to MT include not only maintenance manuals and technical reports - which in many cases appear to be ideally suited to the technology - but policy papers, administrative documents and even journal articles.

In most cases, translation quality for dissemination needs to be high and in some cases it needs to be excellent. Here, machine translation can often be used as a basis for human editing up to the required acceptable standard. Particularly when texts are repetitive and rich in technical terminology, machine translation can be a useful aid in reaching top quality standards.

By far the most common and successful application of machine translation for dissemination or publication purposes is indeed the translation of maintenance manuals. Most MT systems, both large and small, are being used in this way. Large corporations such as Xerox, IBM and Siemens have already achieved quite a record of success, while small hardware and software suppliers are now beginning to report encouraging results with desktop MT software.

The best results here involve a combination of careful source document preparation, a dependable level of technical terminology in the MT system, and human post-editing. The major advantages are not just speed and cost but consistency of terminology which provides for more immediate intelligibility.

In the public sector too, institutions such as NATO, some of the UN agencies and, of course, the European Commission itself are also making use of MT to translate technical reports, administrative documents and minutes of meetings. Raw MT quality is sometimes adequate for user requirements and in many cases rapid post-editing (at a rate of say four pages per hour) provides acceptable results. Post-editing is normally carried out by translators but there is increasing evidence that engineers or other subject-field experts can also produce good results.

Finally, use of machine translation via public networking facilities is beginning to have a considerable impact. In France, it is already being used in significant volumes on the Minitel network where Gachot S.A. provides a number of on-line services using the Systran system. In Canada, the Smart system is being used by the Department of Employment to translate job descriptions between English and French for coast-to-coast access. In Europe, experiments are already underway to combine multilingual database interrogation packages with machine translation in order to provide the non-specialist with rapid and reliable means of accessing foreign language databases.

What remains to be done?

Machine translation can hardly be regarded as a technology in its own right. For it to be used successfully by the non-expert, much remains to be done to overcome many of the technical problems which often outweigh its advantages.

On the one hand, there is the problem of document preparation. The non-expert user sitting at his PC or Minitel terminal knows nothing of the workings of the translation software. He is unaware of the fact that a spelling error, missing punctuation or non-standard formatting will lead to translation errors.

Here progress can be made at two levels. On the one hand, spelling correction technology can be integrated in the automatic interface to the MT system while on the other, a degree of online screen editing can be introduced to draw the user's attention to syntactic and even semantic problems in his draft. This type of technology is developing quickly but improvements in user-friendliness are called for.

In addition, as companies with large multinational requirements become more aware of the cost of translation activities (which can extend to 10% of production costs), it is probable that they will pay more attention than in the past to document drafting. The editing or critique software packages now on the market are desgined to discipline authors and their secretaries in the use of vocabulary and syntax in order to reduce to a minimum the possible ambiguities in a source text. This approach makes not only for better comprehension in the source language itself but for quicker and more reliable translations. Above all, source texts drafted along these lines are far more suitable for machine translation than undisciplined drafts.

Several companies have already adopted this strategy, particularly in connection with maintenance manuals. Extensions to other types of document, for example report writing, can be expected to follow soon.

Current trends

Over the past year, we have seen a number of encouraging extensions to the machine translation market. Logos, Metal, Systran and Tovna have all been successful in finding new customers while sales for desktop packages such as those supplied by Linguistic Products also appears to be on the rise. Extensions to new language pairs have kept pace with the applications side although now, as in the past, there has been a tendency to oversell all new extensions and developments.

Some MT packages, though, have been the victims of restructuring or new company policy. Alps, who still support their computer-assisted translation packages, have concentrated their efforts on translation services in general, particularly through the acquisition and networking of a number of large translation bureaux. Weidner, which had a number of MT packages for European language pairs on PCs, appears to have discontinued reliable support after being taken over by the Japanese company Bravice. Bravice itself, on the other hand, seems to be making considerable progress with English-Japanese and Japanese-English versions of the software.

The Canadian MT market, in particular, appears to be expanding. Logos, Smart and Tovna all have applications there for English-French, mainly in connection with translation projects supported by government funding. However, the ambitious four-million-dollar Gigatext project supported by Saskatchevan seems to have run into serious difficulties.

Systran has been used more extensively by NATO, Xerox, the US Air Force and on the Gachot Minitel network. The European Commission has brought the system on line for internal users (25,000 pages translated in 1989) and is embarking on major applications of the software for the translation of patent literature in collaboration with the European Patent Office.

Last but not least, the Japanese giants who nearly all have MT developments have continued to make progress on the applications side. Several systems are now operational for English-Japanese and Japanese-English although hard statistics on actual users are difficult to obtain.

Progress on MT research

Over the past few years there has been a steady increase in the MT research sector. As we have already seen, the most notable developments have been in Japan where all the large computer manufacturers are developing systems for English-Japanese and Japanese-English and to some extent for other language combinations. The most successful to date appears to be Fujitsu with its Atlas systems.

In Europe, the major research project continues to be Eurotra cofinanced by the European Community and its Member States. It was originally hoped that pilot systems for all the European languages would become operational by the end of 1990 but this goal is proving more and more difficult to achieve. Eurotra objectives for the future are likely to be based more on providing a range of language-processing products for the various EC Member States than on MT alone.

Other projects in Europe include DLT (Distributed Language Processing) in the Netherlands, which is based on the use of Esperanto as a pivot language, and Rosetta - supported by Philips in the Netherlands - which is expected to produce the first operational results in 1990 in systems translating between English, Dutch and Spanish.

In the United States, IBM has once again become involved in MT development, mainly for the translation of its own technical documentation. A number of European universities and research centres are involved in their LMT (Logic-programming-based Machine Translation) project with the development of prototype versions covering English, Danish, French, German and Spanish.

One of the developments which could provide interesting results in the medium term is the Carnegie-Mellon Knowledge-Based Machine Translation project. As its name implies, the project is aimed at using artificial intelligence to resolve natural language ambiguities. As the cost of such developments is very high, even for a narrow subject area, the project could well run into financial difficulties. The approach itself is, however, quite an interesting one.

By and large, though, MT research results have been rather disappointing. Some large projects such as Calliope in France have been terminated. The Japanese systems have proved more difficult to develop than originally anticipated and Eurotra has suffered from difficulty in coordinating developments in the various countries concerned.

With the possible exception of Tovna, the result has been that more traditional approaches to MT have been generally more successful than innovative strategies.

Selection of a system

In my introduction I pointed out that the most important aspect of a typology of machine translation was to assist the user. I have now given an overview of current developments and prospects for the future but perhaps for many it is not a very good basis for choosing an MT system for practical application.

One of the key questions is, of course, "Are you principally concerned with publishing information or with collecting information?"

If you need to publish information, you are probably already employing translators (either in-house or under contract) to ensure that your quality requirements are met. If you decide to turn to machine translation, you will no doubt wish to maintain similar standards.

The criteria you should look at most closely in choosing an MT system can be summarized as follows:

- Has the system already been developed for the languages and subject areas which are of interest to you?
- Can the supplier provide names and addresses of users who have suffient experience of the system to discuss its merits?
- What additional developments (if any) will be necessary to bring the system up to the quality you require (at whose cost and over how many months)?
- How easy will it be to integrate the system into your own existing technical infrastructure?
- Can you take action to improve the quality of your source-language documents (particularly important if more than one target language is required)?
- What measures can you take to ensure that post-editors will indeed be able and willing to make efficient use of the system?

As I may have implied, the cost of a system (whether under a purchasing or licensing agreement) may not be the key factor. Most users have found that integration and further development costs - particularly on dictionaries - are likely to cost far more than the initial installation. In addition, it might well prove difficult to convince translators that they really have something to gain from the use of an MT system; they might well be opposed to changing working methods or becoming a "slave" to the machine. User-friendliness, particularly as far as post-editing is concerned, is thus of the utmost importance.

If you are primarily interested in collecting or scanning foreign language information, then your priorities are likely to be rather different. These might be:

- Can the system deal with a wide range of text types and subject fields?
- Is the quality of the output (for your language pairs) readily intelligible without human intervention?

- Is the supplier likely to provide new improved versions which will increase the level of performance you need? (As a user of raw output, you are less likely to be willing or able to participate in system improvement than the "publishers".)

- How fast is the turnaround of the MT system?

- Does the amount of material you need to scan justify the investment?

Both groups of users would also be well advised to look into ways and means of installing suitable peripheral equipment to be used in connection with machine translation. This might include:

- optical character reading for inputting hard copy;

- sophisticated word processing software for text preparation and any pre- or post-editing;

- grammar and style checkers;

- suitable telecommunications facilities (if required).

The future

Over the next ten years, machine translation is likely to be used more and more extensively, particularly for many routine types of translation processing as well as for information assimilation purposes. Technical documentation, which is already by far the largest source of translation, will increasingly be submitted to MT processing as the drafting of source material improves.

We are unlikely to see any really revolutionary approaches to MT processing. Existing systems will continue to improve with experience and new developments will tend to fall back on well-established processes as the difficulties of programming new linguistic strategies are encountered in practice.

The main users will be multinational corporations and international organizations; database suppliers and all those involved in the on-line information industry will also become dependent on machine translation as the largely English-language information resources come into multilingual access and use.

By the year 2000, Japan is likely to be the main supplier of MT systems and services. Europe will continue to make use of its linguistic heritage in extending and improving projects originating in the United States and Japan but it is questionable whether it will be successful in developing any major systems of its own.

Systems will become more user friendly as improved peripherals are introduced whether on stand-alone systems running on PCs or as a means of improving access to larger systems via telecommunications. Whatever the approach, standardization of document architecture, telecommunications protocols and natural language character sets can be expected to pave the way for increased integration between MT systems and peripheral software in general.

Input technology will also have a major impact on MT use as optical character reading improves and voice technology develops.

Finally, typology itself is likely to evolve once more as market forces compete on two basic fronts: integrated desktop software on ever more powerful machines versus machine translation services provided by telecommications from remote, but ever more sophisticated hosts.

Whether or not it will be possible to carry out machine interpretation between various languages as voice analysis techniques are developed for automatic dictation still remains a largely unanswerable question. Expectations are high, particularly in Japan, but developments - as in traditional MT - are taking longer than expected.

L'environnement technique de la traduction assistée par ordinateur.

Albert YANEZ
Conseiller du Directeur
CEDOCAR
26,BD. Victor
75996, Paris-Armées.

Résumé:

Paradoxalement, alors que les besoins déclarés ou potentiels de traduction et d'interprétation sont énormes et de plus en plus pressants, alors que des progrès spectaculaires sont réalisés dans l'ensemble du secteur des technologies de l'information, la TAO semble marquer le pas, et paraît même en régression dans certains pays.

La raison première en est que cet environnement technique est trop souvent déplorable, sous tous ses aspects et notamment sous l'angle de l'interface homme-machine et ergonomie des systèmes.

Pour la clarté de l'exposé on examine successivement:

-la manière dont le problème se pose aujourd'hui: rappel des enjeux et des différents types de besoins et domaines d'application pouvant conduire à des environnements différents.

-l'environnement technique dans la phase recherche et développement, où l'on distingue l'informatique et aspects connexes d'une part, la linguistique d'autre part, cette étape conduisant à l'étape d'industrialisation du produit ou d'une version du produit.

-l'environnement opérationnel où le système, encore bien fragile et critiquable, considéré cependant comme "défini perfectible", a besoin de s'intégrer dans une application ou chez un client. Cette intégration sera possible si un certain nombre de conditions sont remplies, et notamment si dans les faits le traducteur est réellement aidé. Celui-ci peut apporter beaucoup dans la vie opérationnelle du système si l'ergonomie, pour ce qui le concerne, est appropriée, grâce par exemple à un découpage judicieux des tâches, afin qu'il conserve celles qui sont normalement de son ressort, et afin qu'il conserve la responsabilité de l'"oeuvre seconde" que constitue la traduction définitive.

La conclusion est en forme d'une série de recommandations qui résument les points sur lesquels il faut être attentif si l'on veut améliorer les systèmes actuels et obtenir une acceptation plus grande et une performance accrue des systèmes futurs.

oo0oo

Les enseignements de l'histoire de la TAO: Un serpent de mer?

Il importe de garder en mémoire que l'histoire déja longue et chargée de la traduction par machine est faite d'une suite de proclamations excessives optimistes ou euphoriques alternant avec des périodes de silences et d'oubli, un peu comme il en va en matière d'OVNIs (objets volants non identifiés). Un tel parcours a eu pour effet d'entamer la crédibilité des utilisateurs potentiels aussi bien que des organismes qui finançaient les recherches. La figure 1 présente les points-clés de cette histoire. Il est interessant de noter qu'on s'est arraché à Munich (Summit II - Août 1989) le rapport de la JEIDA (Japan Electronics Industry Development Association) intitulé "vision japonaise de la TAO à la lumière des considérations et des recommandations du rapport ALPAC". Les japonais, tirant à leur manière eux aussi les leçons de l'histoire, prennent aujourd'hui le contre-pied du rapport ALPAC (automatic language processing advisory committee). Le rapport de la JEIDA se fonde sur l'énorme marché de la traduction au Japon pour recommander des investissements massifs dans ce secteur.

Il est intéressant de noter aussi, dans ce rappel historique, que le langage Prolog, qui avait été inventé en France pour des besoins de traduction automatique a été finalement adopté par le Japon dans les projets liés à l'intelligence artificielle et s'est généralisé dans diverses applications.

Elimination des traducteurs?

Le concept de traduction automatique qui sous-entendait que la machine allait apporter la solution a fait place progressivement à la notion beaucoup plus réaliste et plus modeste de traduction assistée par ordinateur, où l'on reconnait avec un peu plus d'humilité que l'objectif sera moins ambitieux et que la solution ne pourra être issue que de la conjugaison des efforts des linguistes et informaticiens, chercheurs et promoteurs de systèmes d'une part, des utilisateurs, notamment chefs d'entreprises et traducteurs d'autre part dont on ne peut se passer pour la phase de développement d'un produit qui restera toujours perfectible et donc dépendant des traducteurs.

Problèmes sous-estimés.

Ces derniers d'ailleurs étaient restés très sceptiques sur les résultats à escompter d'une traduction automatique qui résoudrait avec une logique binaire les problèmes tout en nuances auxquels ils sont confrontés, problèmes liés par exemple aux figures de style (voir ci-dessous) beaucoup plus fréquentes qu'on ne le pense , même dans la langue technique, et aux maladresses des auteurs qu'ils ont souvent à aider dans des démonstrations maladroites exprimées dans un jargon obscur ou ambigu ("the fish found dead in the river will be replaced by farmers"). Dans un de ses essais, Eugene Garfield voyait même dans la rédaction une fonction à laisser à des spécialistes (a job for professionals).

LES FIGURES DE STYLE.

la synecdoque: la partie pour le tout et vice versa. (faire de la voile)

l'anacoluthe: changement brutal mais licite de construction grammaticale

l'antonomase: emploi d'un nom propre au lieu d'un nom commun. (Wall Street n'a pas réagi.)

l'image: la charrue avant les boeufs. Etoile rouge (les soviets) sur la grande bleue (méditerranée).

l'analogie: cette affaire est un serpent de mer. On nous mène en bateau.

l'ellipse: omission des mots qui ne sont pas indispensables: train rentré (pour train d'atterrissage)

la métaphore: transfert de signification (brûler de désir).

la litote (understatement): il s'est éteint, pour "il est mort".

les archaïsmes, les néologismes non encore homologués, les jeux de mots, les proverbes...et tutti quanti, toutes les expressions et idiotismes qui ne sont pas forcément faciles à déceler (trop c'est trop, pour enough is enough) et puis les faux-amis (I recognize you),... et les homographes, et la polysémie, et les abreviations, symboles, codes, sigles ou formules de plus en plus fréquents notamment dans la langue technique...On voit à quel point le parcours est semé d'embûches redoutables. Et ceci sous-tend qu'au lieu d'ignorer les traducteurs, on aura besoin d'eux parce qu'ils connaissent bien les pièges à déjouer entre langue source et langue cible. N'oublions pas par exemple que Peter Toma, le père de Systran, était d'abord polyglotte.

Téléscopage de la phase développement et de l'industrialisation.

Cet empressement des chercheurs à annoncer des résultats et des succès a masqué longtemps le fait que la recherche ne pouvait pas être directement suivie de l'application, et qu'il fallait nécessairement passer par une longue et rude étape de développement, d'apprentissage, avec, en désespoir de cause, le concours de traducteurs connaissant bien la langue source et la langue cible, puis par une phase d'industrialisation pour aboutir par exemple à un produit portable, compatible avec les ordinateurs les plus couramment utilisés, permettant une utilisation interactive avec une bonne ergonomie, des temps de traitement acceptables, une prise en compte immédiate des observations des utilisateurs. Sur tous ces points les progrès ont été et sont encore très lents et incertains et pourtant le véritable succès repose pour beaucoup sur ces aspects trop souvent négligés. L'utilisateur croyait pouvoir obtenir un produit clé en main; il a été surpris de constater que c'était à lui qu'il incombait de nourrir l'enveloppe qui lui était remise.

Ainsi l'histoire de la TAO a permis de mettre en évidence un certain nombre de points qu'il faudra désormais avoir soin de garder à l'esprit, par exemple le fait qu'il ne faut pas placer trop haut les objectifs et qu'il faut si possible choisir des domaines d'application bien limités et circonscrits, et ne pas demander à un système de traduire n'importe quel document.

La réussite de TAUM METEO, tout à fait opérationnel et rentable, en témoigne (coût: 0,03 dollar canadien par mot, pour un débit de 3,5 millions de mots par an.(la figure 2 montre qu'il s'agit là d'un problème relativement simple, si simple que dans ce cas on peut effectivement parler de traduction automatique, puisqu'aucune révision n'est nécessaire)..

La notion d'étapes distinctes

L'histoire nous apprend aussi qu'il faut séparer grammaires et mécanismes ou algorithmes d'analyse d'une part, dictionnaires et outils terminologiques d'autre part, de façon à faciliter l'évolution en intégrant plus facilement les progrès réalisés dans chacun de ces domaines.

Il faut aussi considérer que la qualité est liée aux développements à plusieurs niveaux:

1.Translittération, vérification et préparation du texte (tout ce qui peut entrer dans la phase dite de "pré-édition" qui est l'ensemble des taches permettant à la machine de savoir reconnaître au mieux ce qui lui est présenté..

2.traduction mot à mot (à partir de dictionnaires plus ou moins évolués, pouvant aller jusqu'à une "navigation" dans une base de connaissances terminologiques organisée, de type thésaurus de descripteurs par exemple (base qui peut être organisée soit a priori soit à partir du corpus entré).

3.analyse syntaxique (arbres syntaxiques permettant d'aller au-delà du simple mot à mot, en identifiant sujet, verbe, compléments...C'est le niveau où l'on sait reconnaître la construction de la phrase, indépendamment de la reconnaissance de son contenu informationnel.

4. Le quatrième niveau, qui vient s'ajouter à l'analyse syntaxique, est celui de l'analyse sémantique. "He is a gas" ne devient compréhensible que si He est rapproché de hélium. En japonais notamment, où l'ordre des mots n'est pas rigide comme en anglais, une analyse purement grammaticale laisse subsister bon nombre d'ambiguités (1).

5. Enfin et surtout la qualité n'a de chances d'être atteinte que si l'on dispose d'indicateurs de contexte, ce qui suppose que la machine ait une connaissance du monde extérieur, une certaine faculté de raisonnement à partir des faits ou des données qui lui ont été fournis. Une bonne part des vicissitudes de la TAO vient du fait que jusqu'ici ces indicateurs

de contexte étaient pratiquement inexistants, et le grand progrès viendra du recours aujourd'hui possible à l'intelligence artificielle.

Rapprochement avec les systèmes experts.

La tendance serait à organiser une communication entre les mécanismes de traduction et la base de connaissances linguistiques, cette communication étant gérée par le KBMS (knowledge base management system), cette gestion impliquant un retour d'information servant à l'accroissement et à l'amélioration des connaissances au cours de la vie du système (2), autrement dit on tend à se rapprocher de la philosophie des systèmes experts: règles utilisées par un moteur d'inférence et le recours à une base de connaissances, dont le contenu est géré par des cogniticiens tirant le meilleur parti possible du savoir faire d'experts qui ici pourraient bien être les traducteurs et les interprètes de conférence, qui sont les véritables experts et dont le concours, répétons le, permettra d'éviter les bourdes encore trop souvent rencontrées dans les résultats après des années de recherche et de développement!

Impact du marché sur l'environnement

Perception de l'enjeu.

L'enjeu est devenu beaucoup plus important. Il est surtout mieux perçu et pris en compte au niveau politique.
L'investissement dans la TAO est reconnu comme une nécessité parce qu'on s'accorde aujourd'hui à reconnaître l'importance de l'investissement immatériel à côté de l'investissement matériel, ceci non seulement dans le secteur scientifique et technique, mais dans le monde des affaires, assurances, banque, tourisme, droit, religion, bref de tout ce qui touche à la culture et à la communication entre les peuples. On découvre que la réduction de la barrière linguistique est le plus grand défi de cette fin de siècle.
L'Europe, avec une grande sagesse, fait tout pour préserver chaque langue.
"Lorsqu'une langue meurt, avec ses couleurs, ses nuances, le peuple meurt aussi" (Maila Talvio - Finlande- Pensées Eternelles.)

Le babélisme dans le monde.

Or la réalité est qu'il existe environ 3000 langues vivantes dans le monde, parmi lesquelles il faut faire des choix liés aux enjeux culturels et plus prosaiquement aux marchés à escompter.

Leur importance relative peut se mesurer:(3)

-selon l'ethnie: le chinois d'abord, puis l'anglais (8,6%), puis l'hindi, l'espagnol, le russe, le français n'étant que 12éme avec 117 millions de personnes.

-selon l'effectif des locuteurs: l'anglais (30%) loin devant le portugais (7%), le russe (6%)

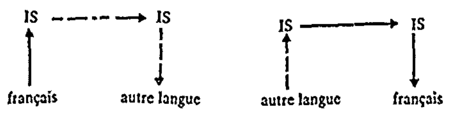
-selon le volume des publications scientifiques et techniques: l'anglais (plus de 50%), le russe, l'allemand, le français et le japonais totalisant 40% supplémentaires

-selon la production littéraire

-selon la qualité des auteurs: 21 prix Nobel attribués à des ouvrages en anglais, 12 pour le français, 9 pour l'allemand...

Politique européenne: Systran puis Eurotra.

L'Europe a renoncé à adopter une langue unique (anglais, français ou espéranto). En 1975 la CEE a acquis Systran, comme début de solution mais, prenant conscience de ses limitations et insuffisances, elle a lancé le programme Eurotra (European Translator). Le modèle à transfert choisi pour Eurotra implique que les modules d'analyse et de génération de chaque langue soient conçus dans une optique monolingue. Chaque nation est en charge de l'analyse de sa langue et du transfert des autres langues vers sa langue (2). On aboutit ainsi à 72 modules de transfert pour les neuf langues officielles de la Communauté. Par exemple l'équipe française est chargée, pour chacune des huit autres langues, du travail indiqué en trait plein dans les diagrammes ci-dessous, où le symbole IS désigne la structure d'interface.



Transformation de l'environnement.

Une telle évolution dans la prise en compte du problème de la TAO a pour conséquence de transformer radicalement l'environnement. Les développements sur Systran sont laissés aux utilisateurs tandis qu'avec Eurotra on mobilise dans chaque nation les équipes universitaires les plus compétentes dans l'analyse de la langue vernaculaire et sa représentation, ceci très globalement et indépendamment de besoins particuliers qui pourraient par la suite interesser tel ou tel utilisateur, dans tel ou tel contexte, celui-ci pouvant alors apporter le complément d'une base de connaissances propre à son environnement spécifique.

On entre ainsi dans un univers tout différent. La TAO devient une composante ou un segment d'application du traitement et de l'industrie de la langue, parmi de nombreuses applications connexes qui vont toutes se renforcer.

La figure 3 montre que désormais le traitement de la langue constitue de plus en plus une discipline en soi s'exerçant au profit de la TAO mais tout aussi bien de la communication en général: reconnaissance du contenu des textes ou du discours, generation des documents par voie électronique en vue de leur traitement, que ce soit dans le cadre de l'informatique documentaire, du dialogue avec des systèmes experts, de l'étude statistique ou conceptuelle ou informationnelle de contenu.

Un exercice intéressant, si l'on veut

confirmation de ce phénomène, est de faire une rapide analyse qui sans aller jusqu'à l'analyse bibliométrique peut être la suivante: à partir du fichier bibliographique Inspec par exemple voir, en utilisant une commande toute simple comme "..memt" de Questel Plus, ce qu'est l'environnement sémantique de la traduction assistée ou automatique d'une part, du traitement de la langue d'autre part:(cf. figure 4).

Pour la traduction on peut noter la place importante des dictionnaires, puis ... des politiques gouvernementales, ce qui est le signe de la prise de conscience dont on parlait plus haut, puis des SGBD, des bases de connaissances, des applications de la microinformatique, du traitement de texte, jusqu'à l'édition assistée et les systèmes experts.

Dans le cas où le traitement de la langue est pris comme point focal, on trouve les dictionnaires, le traitement de texte, la formation assistée par ordinateur, l'édition électronique, l'indexation, les applications de la microinformatique...

Regroupements autour du traitement de la langue.

De même que l'on verra des équipes universitaires jusqu'ici dispersées se regrouper, pour se partager les tâches au lieu de s'ignorer ou de se concurrencer, on verra s'opérer dans les entreprises des regroupements permettant de réunir tout ce qui est connexe et interdépendant, qu'il s'agisse par exemple de l'intégration de la chaîne de production des documents, depuis l'aide à la rédaction jusqu'à la diffusion en passant par la traduction, par la normalisation et bien entendu par des regroupements de tâches, de compétences et de métiers.

L'industrie de la langue ou le traitement de la langue apparaît véritablement comme un nouveau paradigme. G.Dosi a défini le paradigme technologique comme un ensemble de problèmes, de procédures et de tâches liés au développement technologique, dans lequel les forces du marché et la demande vont agir comme un mécanisme de sélection (5). Une fois qu'une voie de changement technique a été créée, celle-ci a une dynamique propre, qui définit les directions dans lesquelles l'activité de résolution du problème se déplace. On passe ainsi désormais d'une conception systémique (la TAO) à un ensemble de besoins de fonctionnalités (placées dans l'environnement du traitement de la langue et finissant par s'intégrer).(6)

Consolidation et importance de l'industrie de la langue:

Le marché va donc se trouver consolidé sous l'effet de plusieurs facteurs liés aux progrès technologiques et aux avancées du génie logiciel et des linguiciels TAO mais aussi en amont et en aval de la TAO.

Par exemple le secteur de l'édition, ou tout au moins ceux des éditeurs qui se sont engagés dans la même voie, vont bénéficier de la possibilité de rayonnement accrue qu'apporte la TAO intégrée à une chaîne d'édition. Dès à présent les brevets japonais par exemple ne sont plus un champ clos. Par le truchement des banques de données bibliographiques, lorsque l'édition n'a été prévue que dans une seule langue, l'information sur l'existence et le contenu des documents publiés dans cette langue se trouve accessible par des utilisateurs d'autres langues, soit que les producteurs de ces banques de données aient entré par exemple un thésaurus multilingue, comme dans le cas du fichier Pascal de l'INIST (spécimen fig. 5), qui se trouve indexé en trois langues et ainsi accessible lorsque les questions sont posées en français, en anglais ou en espagnol, soit que le serveur ait acquis un logiciel qui permet cette transposition d'une langue vers l'autre même si le texte n'a pas été préalablement indexé par des mots-clés. Cette transposition pourra être intégrale ou partielle, accompagnée de la mise en évidence du contenu essentiel du document, ou pourra focaliser sur les aspects en relation avec l'intérêt ou le "profil" de l'utilisateur, de sorte que d'un même coup l'on va pouvoir détecter l'information utile de manière beaucoup plus fine que par les seuls opérateurs logiques appliqués brutalement entre des mots, ceci en passant par des analyseurs de texte utiles aussi bien pour la compréhension de la question et éventuellement un dialogue avec l'utilisateur, que pour la sélection et les transpositions utiles dans la langue de l'utilisateur.

D'une manière générale les banques de données en ligne constituent une source importante d'information linguistique et une aide à la traduction, ainsi que l'explique Hikomaro Sano (19).

Une autre possibilité est de placer, au niveau de l'utilisateur, un dispositif permettant de transférer les résultats d'une interrogation en ligne sur un serveur de traductions, du type Systran, et recueillir une traduction brute qui, dans certains cas, pourra se passer de post-édition (par exemple s'il s'agit de parcourir titres et résumés signalétiques de documents dont on veut vérifier le contenu avant de les commander ou de les faire traduire.

On peut aussi imaginer la traduction par machine a priori de l'ensemble de la base bibliographique et de ses mises à jour. Même si cette traduction est imparfaite, le spécialiste qui lira les titres et résumés n'aura pas trop de peine à apporter, presque inconsciemment, les corrections utiles. Personnellement j'ai pu constater qu'un résumé en allemand traduit par machine en français permet d'appréhender le contenu de manière suffisante pour déterminer si le document méritera ou non d'être commandé pour être traduit.

On voit donc que c'est tout le marché de l'information en ligne qui va ainsi pouvoir s'agréger au marché de la TAO. Or le chiffre d'affaires des services en ligne est déjà de $ 4 à 7 milliards(71%) aux Etats-Unis, de $ 1 à 1,5 milliards (18%) en Europe, $0,5 milliard (10%) au Japon... et de moins de $ 0,05 milliard (1%) dans le reste du monde.

Des banques de données jusqu'ici

totalement hermétiques ou de peu d'interêt économique voient leur facteur d'impact augmenté, ainsi qu'est renforcé l'impact des produits ou services et de la culture qu'elles véhiculent.

Parallèlement aux systèmes serveurs de traduction utilisables aussi bien par le grand public, par minitel, que par des organisations dotées de moyens performants spécialement étudiés, on a vu se développer, avec la mini et la micro-informatique le marché de petits systèmes qui peuvent être très efficaces s'ils travaillent dans un domaine bien circonscrit, avec un vocabulaire bien maîtrisé, même si c'est avec une syntaxe excessivement simple. C'est le cas de ALPS, de Macrocat (Wiedner) ou de Bravice (Japon). Leur succès s'explique aussi par le fait qu'il existe dans les organisations une information confidentielle qui ne peut être envoyée sur un serveur extérieur. Il faudra donc que les serveurs de TAO pensent à des versions micro de leurs logiciels, tout comme Questel a produit micro-Questel, versions à implanter dans les entreprises, si ces serveurs detraduction ne veulent pas s'exposer à perdre une part de marché.

Du point de vue des systèmes, la répartition en Europe était récemment la suivante: Logos(26%), Weidner(23%), Ericsson(16%), Systran(13%), Alps (12%) - (5)

La CEE consacre une part importante de son budget à la traduction, soit environ 1 milliard de francs par an, et elle occupe 1800 traducteurs répartis entre Bruxelles et Luxembourg. Son choix en faveur de Systran, dont elle a en grande partie financé le développement (4 MECUS de 1977 à 1982) a été contesté, mais une intelligibilité suffisante de la traduction brute a tout de même été atteinte, tout comme à l'OTAN (specimens in fine). Depuis lors elle investit dans des recherches propres (EUROTRA), tout en restant un des principaux utilisateurs de Systran (6). Le marché mondial est estimé à 3 milliards de dollars par an représentant 150 millions de pages et occupant 175000 personnes. Certains disent que ces chiffres sont bien en deça du marché potentiel qui apparaîtra lorsque des systèmes plus conviviaux et plus performants seront prêts. Toujours est-il que la demande devrait s'accroître de 50% en cinq ans, et que la TAO devrait occuper rapidement 5 à 15% de ce marché. Ces chiffres sont difficiles à vérifier. On ne peut que faire des recoupements entre renseignements de diverses provenances.

En matière de recherche, le Japon a prévu un gigantesque effort national en faveur de la TAO où devrait s'engloutir, dans les 12 prochaines années un budget comparable à celui de l'ICOT pour les ordinateurs de 5° génération. Deux importants programmes de recherche de plusieurs milliards de francs ont démarré: l'un pour la création de l'Electronic Dictionary Research Institute (1,5 milliards de francs, avec huit industriels, l'autre à l'initiative du Ministère des Postes, pour la mise au point d'un téléphone traducteur (4 milliards de francs). En 1985 on recensait déjà au Japon 18 projets de TAO, et plusieurs systèmes de 2° génération, avec approche sémantique de modèles de langage sont déjà commercialisés. La société NOVA propose notamment pour le couple anglais japonais une station de travail qui traduit en une heure 50 pages, soit 20000 mots. Le système vendu au prix de 5,55 millions de yens pourrait être vendu à 200 exemplaires en 1989. ATLAS 2 de Fujitsu et Pivot de NEC peuvent traduire jusqu'à 60000 mots à l'heure.
Ceci n'empêche pas le Japon de travailler aussi sur Systran et même d'obtenir d'importants contrats du gouvernement américain pour la traduction japonais-anglais. Systran traduirait dans ce couple 1,2 million de mots (6000 pages de format A4) en une heure, avec une précision de plus de 85%. Cette décision du gouvernement américain est destinée à améliorer le déséquilibre des échanges d'information entre les Etats-Unis et le Japon. Nous savons aussi que l'université d'Edimbourg coopère avec le Japon sur le traitement de la langue parlée, de même que des chercheurs français apportent un concours dans une direction très voisine, celle de l'interprétation par machine (7). C'est ici que les progrès les plus spectaculaires sont à attendre, avec le développement des machines à " dictée magique" et l'analyse des phonèmes.

Avec le développement du téléphone dans les phases RNIS (ISDN) et POST-RNIS et en parallèle les réseaux neuromimétiques et les machines connexionnistes, on va se trouver dans un environnement informatique et télématique particulièrement adapté enfin au traitement de la langue écrite et parlée.

Cette perspective doit être prise en compte dans toute évaluation de la croissance du marché de la TAO.

A l'autre extrémité de l'éventail du marché, et beaucoup plus modestement, il y a place pour des aides simples et portables, destinées à certaines applications, par exemple chez les militaires et les pilotes en particulier, dans un cadre d'interopérabilité qui doit exister sans accroissement du stress auquel sont déjà soumis les personnels. Très modestement, SANYO propose un dictionnaire électronique portable anglais-japonais de 35000 mots destiné aux étudiants et aux hommes d'affaires. Les possibilités déjà offertes par les disques compacts CD-ROMs et les logiciels hypertexte et multimédia viennent renforcer la probabilité d'éclosion d'applications très diverses, et donc il ne faut pas ignorer cette part de marché, et être attentif à ces niches ou créneaux associés au développement de la synthèse de la parole, et où il y a place pour la traduction : aides aux handicapés, assistance aux opérateurs, jeux électroniques, traducteurs de poche, par ordinateur, notamment enseignement des langues, renseignements téléphoniques et messageries vocales, , contrôle des tâches, alarmes vocales, claviers vocaux, annonces parlées (marée et météo, horaires des trains ou avions, stations d'autobus, synthétiseurs de trafic) ou autres applications du traitement de la parole qui a fait bien des progrès , (machines à dicter ou commande vocale de Crouzet pour le Rafale).

Si les regroupements se font entre

domaines d'application voisins, le marché de la traduction assistée devrait croître très vite. Actuellement la dispersion des outils périphériques (de traitement de texte ou d'édition, vérificateurs orthographiques, lecteurs ou numériseurs) rend difficile l'évaluation de la progression du marché.

Par exemple le marché mondial des industries de la langue parlée, marché qui va lui-même interférer avec celui de la langue écrite, était évalué à 14 millions de francs en 1984 et il devrait atteindre 28 milliards dès 1990 (un tiers pour la synthèse de la parole, deux tiers pour la reconnaissance verbale) soit deux mille fois plus en six ans...

Ainsi si le marché de la TAO n'a pas vraiment décollé en Europe et en Amérique du Nord, on peut s'attendre à une explosion de l'ensemble du marché del'industrie dela langue, explosion contrôlée par les japonais, à moins qu'apparaisse le programme cadre européen attendu comme réplique à la stratégie internationale des japonais. Un représentant de la Commission aurait indiqué à Munich deux hypothèses pour le programme LIFE (Industries de la langue en Europe) de 150 MFF à 2 milliards de FF.)

Environnement technique proprement d..

Ce tour d'horizon général sur l'évolution a montré que la TAO est aujourd'hui sortie de son isolement, qu'elle doit passer du stade du laboratoire à l'industrialisation pour se placer dans un environnement opérationnel, au même titre que d'autres applications du traitement de la langue naturelle, et qu'elle doit pouvoir s'intégrer dans la chaîne de traitement documentaire.

Il faut rappeler que l'environnement technologique s'est fondamentalement transformé depuis les débuts de la TAO. Il me semble que seul Peter Tomá, le père de Systran, avait à l'époque une vision de ce que serait cette évolution: possibilités accrues des mémoires centrales et surtout des mémoires périphériques à des prix de plus en plus faibles, evolution des langages de programmation, apparition de réseaux fiables et à large bande pour la transmission des données, généralisation de stations de travail bureautique et mini ou microinformatique pouvant appeler des serveurs de dictionnaires electroniques , des serveurs de banques de données textuelles et des serveurs de traduction, et enfin apparition de l'intelligence artificielle ; cette évolution technologique rend tout à fait plausible une percée prochaine importante dans le secteur de la TAO ou de l'interprétation assistée par ordinateur, compte tenu des progrès réalisés dans les secteurs connexes de l'industrie de la langue: reconnaissanse de la parole et numérisation des phonèmes, lecture optique..etc.

Integration:

L'intégration dans la chaîne de traitement documentaire peut prendre plusieurs formes: dans le cas relativement bien circonscrit de la documentation technique, par exemple chez les grands constructeurs du secteur aérospatial qui ont à produire et à traduire d'énormes volumes de documentation accompagnant les matériels (à titre d'exemple la traduction de la documentation d'un Airbus demande 60000 heures de travail de traduction, soit 39 années/homme pour un coût de 6 MFF), l'intégration commence au niveau des bureaux d'études avec la CAO, conception assistée par ordinateur, ou avec la fabrication intégrée assistée par ordinateur. Il apparaît de plus en plus discutable et même aberrant (6) de revenir à un support papier encombrant et d'intérêt limité alors que l'information est ou aura été numérisée et balisée par SGML (standard generalised mark-up language) dans le cadre CALS (Computer Aided Acquisition and Logistic Support Initiative), programme amorcé aux Etats-Unis mais déjà suivi par un certain nombre de pays (Eurocals). Ainsi l'information se trouvera accessible en ligne, dans la forme souhaitée par l'utilisateur, et non plus dans une présentation prédéterminée , unique et figée, qui est celle du support papier à partir de l'organisme source, le mieux placé pour la générer et la mettre à jour, directement ou par l'intermédiaire d'une passerelle (gateway). Il est évident que non seulement on arrivera ainsi à des économies substantielles mais qu'on disposera à tout moment d'une information à jour et éventuellement de sa traduction à jour dans la langue des principaux pays clients. Cette perspective n'est pas lointaine dans la mesure ou les analyseurs de texte et autres outils linguistiques utilisés pour la TAO seront de toute façon egalement utiles pour toutes les interfaces telles que interrogation de banques de données multilingues, passerelles pour y accéder, systèmes experts et bases de connaissances associées, et autres applications du traitement de la langue telles que exploration rapide (skimming) et routage systématique vers les utilisateurs.(9)

Site d'implantation de la TAO.

Dans la majorité des cas son site d'implantation idéal et évident pour lui donner les conditions d'environnement les plus favorables sera le service d'information ou de documentation qui existe a divers degrés de développement dans toute organisation. On trouve en effet déjà dans l'activité du service de documentation, qu'il ait ou non une mission précise de traduction, toutes les facettes que l'on a signalées dans l'environnement TAO puisque celui-ci a en charge la génération, la collecte, l'archivage, le traitement, la selection et la diffusion des documents et surtout la gestion de leur contenu informationnel en passant par l'analyse et l'indexation, qui sont également assistés par ordinateur. On y trouve déjà nécessairement du personnel linguiste, puisque l'information traitée est en plusieurs langues, personnel spécialisé de surcroit dans les techniques touchant la couverture du centre. Ce personnel a déjà une longue expérience des problèmes de sémantique. Il a créé et utilise des lexiques et thésaurus monolingues ou multilingues, il interroge des banques de données terminologiques, il développe et utilise des logiciels d'analyse linguistique dans le cadre de la recherche documentaire ou bibliométrique, il manipule dejà lecteurs optiques ou numériseurs, ou traitement de texte pour saisir l'information collectée. C'est là sans aucun doute qu'il faudra renforcer éventuellement les équipes et les moyens et surtout éviter de les dupliquer en les implantant ailleurs.

Les étapes du traitement:

Génération, collecte et saisie du texte, et prétraitements associés à cete phase:

Si l'on veut aller vers l'industrialisation il faut favoriser et encourager la production de texte numérisé, qu'il s'agisse de traitement de texte ou d'édition électronique ou de toute transaction permettant d'avoir au point de départ une représentation du texte qui évite les saisies onéreuses et peu fiables par lesquelles on avait à passer naguère.

Ainsi le texte peut provenir d'une bande magnétique utilisée dans la composition programmée puis débarasseé des signes de composition, ou d'un télédéchargement, ou d'une messagerie électronique. Ce n'est que dans les cas où l'on ne disposera d'aucun support non imprimé qu'il faudra se résoudre à passer par un lecteur optique ou par un numériseur capable d'utiliser des algorithmes de reconnaissance des images et des caractères. Dès la génération du texte il convient d'utiliser au maximum toutes les ressources que la bureautique peut apporter.

Sans aller jusqu'aux contraintes qui ont été acceptées dans le système TITUS (figure 6) développé par l'Institut Textile de France, contraintes qu'on ne peut imaginer que dans un environnement totalement contrôlé, on peut se servir

d'outils et de logiciels tels que ceux que propose la société Microsoft, à partir d'un CD ROM doté des fonctions suivantes:

-dictionnaire de200000 termes(American Heritage)
-dictionnaire des synonymes Roget's
-citations de Bartlett's familiar quotations
-World almanach of books and facts
-ouvrage de rérérence sur l'art d'écrire (Chicago manual of style)
-correcteur orthographique fonctionnant sur un algorithme phonétique
-correcteur d'usage (en fonction du contexte)
-formulaire et lettres-type...

Très vite on entre ainsi dans un processus de pré-traitement et dans le domaine des linguiciels où des aides diverses existent. Par exemple les travaux de Janine Gallais Hamonno ont montré que la syntaxe utilisée par les différentes professions ou communautés scientifiques varie en fonction des spécialités. Par exemple les outils LIDIA (10), paramétrés en français et en anglais, permettent d'améliorer le texte dans une spécialité. L'objectif est de traduire comme si le texte avait été écrit par un spécialiste de la langue cible, c'est-à-dire de ne plus traduire le texte tel quel mais de le modifier pour tenir compte des modes de pensée, de culture et d'expression de la langue cible. Toujours dans le même environnement, ANAGOGE permet de constituer automatiquement des dictionnaires de concepts, de regrouper les champs sémantiques associés par les spécialistes à chacun des concepts, d'analyser les trames rhétoriques des textes entrés et de constituer des bibliothèques de trames dont chacune correspond à un mode de présentation d'un document ou à un type d'argumentation, de connaître les concepts utilisés dans les textes et donc de prendre leur traduction dans un dictionnaire d'équivalences qui évite les erreurs rencontrées lorsqu'on ne s'interesse qu'aux occurences de mots et non aux concepts.

Dans le même ordre d'idée HIERARCHIE permet d'extraire les concepts, d'analyser leur hiérarchie et de créer automatiquement des thésaurus de termes ou expressions désignant les concepts. Il fonctionne pour le français et l'anglais et peut être utilisé pour l'indexation automatique des textes en anglais ou en français, l'alimentation automatique d'une base de données, le routage des messages, l'analyse automatique des traductions (repérage automatique des erreurs, ceci après un paramétrage sur quelques centaines de pages de texte d'une spécialité. Ceci implique que la traduction porte sur des documents d'un volume relativement important, constitués en séries si possible. Ces logiciels sont en cours d'adaptation au russe et au japonais.

La reconnaissance de format, ou la mise au format, fait partie aussi de cette étape préliminaire de "pré-édition". Par exemple, lorsqu'il s'agit de traduire des notices bibliographiques provenant de banques de données, on pourra choisir de ne prendre en compte pour la traduction que le champ titre, le champ résumé et le champ mots-clés, donc de reconnaître les autres champs afin de les ignorer momentanément.

La translittération est aussi une opération amont qui peut être entièrement automatique.

Dans les autres points de détail, la préparation du texte en vue de sa reconnaissance optimale peut être guidée par le système, qui posera des questions sur ce qui lui paraît ambigu ou sur ce qu'il ne sait pas interpréter en première lecture. Il faudra apporter les marques particulières qui renseignent sur chacun de ces points, ou indiquer que l'on peut ignorer tel ou tel obstacle.

Cette opération sera peut- être fastidieuse pour un traducteur, alors que celui-ci sera en revanche indispensable dans la post-édition. Néanmoins ce travail doit être confié à un personnel ayant une certaine connaissance de la langue source et de la langue cible, et qui soit capable de conduire une fonction d'enrichissement ou d'apprentissage pour tout ce qui a un caractère recurrent, ceci en mode interactif si possible.

Terminologie.

L'organisation des ressources en terminologie conditionne la qualité des résultats. Dans le secteur d'activité ou de spécialité de l'utilisateur quel qu'il soit on trouve des termes généraux, constituant un vocabulaire livré en général par le fournisseur du système, par exemple Robert & Collins pour Wiedner, et des dictionnaires sectoriels dont certains peuvent aussi être livrés avec le système. On affectera une priorité à tel ou tel dictionnaire sectoriel en fonction du contenu du document traité. Mais cette ressource terminologique est insuffisante. Il demeure que l'utilisateur doit faire un effort assez important s'il veut maîtriser convenablement la sémantique pour atteindre un niveau suffisant d'intelligibilité de la traduction brute et faciliter la post-édition. Tous ceux qui sont parvenus à des résultats tangibles en matière de TAO ont compris qu'il fallait préalablement passer par cet effort qui constitue un investissement. C'est ainsi qu'ont procédé les Communautés Européennes, l'Aérospatiale et l'Institut Textile de France par exemple.

Il est interessant d'examiner de plus près ce qui se passe à l'Aérospatiale (15). L'ensemble des activités de terminologie est soigneusement coordonné selon le schéma ci-après qui constitue le véritable réseau terminologique de la société, et dont le coordonnateur est en interaction avec les banques de données de terminologie et les dictionnaires produits par la société. L'effort dans ce domaine d'activité est remarquable:

1971 dictionnaire français-anglais 16000 termes

1978 dictionnaire trilingue français anglais allemand.25000 termes

1984 dictionnaire quadrilingue français anglais allemand espagnol 50000 termes

et ainsi de suite avec un dictionnaire des abréviations, un dictionnaire des définitions, et l'accès à une banque de terminologie interne, toutes ces sources étant bien entendu numérisées, et à des banques de terminologie exterieures, nationales, européennes ou internationales.

C'est d'ailleurs probablement grâce à cet investissement dans la terminologie que l'Aérospatiale s'est trouvée en position très favorable pour négocier avec la CCE dès mars 1982 l'utilisation de Systran compte tenu de son apport d'un dictionnaire quadrilingue couvrant bien le secteur aérospatial.

C'est dans ce contexte que l'Aérospatiale peut afficher des coûts globaux de 0,45FF /mot sur les couples français anglais et anglais français, coûts calculés sur l'ensemble des traitements: préparation du texte et lecture optique, traduction brute et post édition affinée. On trouve dans ce même article de l'Aérospatiale (15) des indications détaillées sur les coûts.

Lorsque les dictionnaires sont réalisés en interne, le résultat a toutes les chances d'être bien meilleur, surtout si ce travail est piloté par le service d'information ou de documentation où l'on á depuis longtemps l'habitude de raisonner sur les concepts et non sur les mots isolés ou "unitermes". Tous les spécialistes de l'information savent combien il est dangereux de séparer des termes en relation paradigmatique. Par exemple rayonnement de freinage est une entité tout comme son équivalent bremsstrahlung. On se souvient des échecs qu'avait connu Taube avec ses unitermes et des succès de Mooers avec ses descripteurs. L'atterrissage sur le ventre d'un Tupolev ne peut en aucune manière, si ce point est bien compris, devenir dans la traduction l'atterrissage - sur le ventre d'un Tupolev. En anglais par exemple belly-landing forme un tout, alors que le français permet ce type d'erreur si l'on n'a pas raisonné au niveau conceptuel.

La traduction proprement dite.

Les systèmes dits de 1ère génération utilisaient la méthode directe de passage de la langue source à la langue cible, c'est-à-dire la traduction mot à mot, suivie par une procédure visant à réarranger les mots dans la phrase en utilisant des règles de reconstruction pour aboutir à des phrases acceptables, en dehors de toute compréhension du contenu. On ne peut bien entendu trouver des règles suffisamment générales qui soient applicables à tous les types de contenu de texte, ou alors il faut limiter ces types de contenu comme dans Titus. On est alors passé à une approche interlinguale dans laquelle on s'efforce de comprendre la signification de la phrase à travers une analyse conduisant à une représentation indépendante de la langue cible.

On est passé ainsi de la notion de langage pivot à la notion de structure de transfert ou d'interface que l'on trouve dans Eurotra. C'est cette évolution que suit aussi le GETA de Grenoble à partir d'Ariane-78 et après que ce logiciel ait pu être testé sur un certain nombre de couples de langues(16). Aujourd'hui donc la plupart des systèmes se considèrent comme de 2me ...ou de 3ème génération selon qu'ils font appel à cette structure de transfert et à l'intelligence artificielle et aux règles qui ont pu être expérimentées avec succès, mais dans des espaces limités, avec les systèmes experts. Par voie de conséquence également on a de moins en moins recours au système simpliste que constituent les équivalences proposées dans les dictionnaires, et l'on s'oriente vers une prise en compte plus systématique de "descripteurs" et d'expressions toutes faites.

Certaines langues font aujourd'hui l'objet d'études très poussées visant à la représentation par des graphes de toutes les significations des mots dans tous les contextes d'utilisation en testant les résultats sur des corpus mis à la disposition des linguistes, comme le "Trésor de la langue française"

Le LADL (laboratoire d'automatique documentaire et de linguistique) du Professeur Maurice Gros a mis sur bande 130000 formes verbales du français. Son dictionnaire électronique a pour objet la description de la langue de fond en comble, en n'oubliant aucune expression, aucun idiome. La langue devient ainsi une matière première industrielle utilisable dans toutes ses applications, y compris la TAO.

Selon M. Oreja, secrétaire général du Conseil de l'Europe, les langues qui ne s'industrialiseront pas cesseront d'être des langues véhiculaires, des langues de civilisation.

Le projet Eurolexic, coordonné par la France, est de créer une grammaire et un dictionnaire électroniques d'abord en quatre langues -anglais, espagnol, français, italien - les autres langues européennes devant suivre.

On constitue alors dans chaque langue, et dans le cadre général des industries de la langue, une base de connaissances linguistiques, qui n'est pas destinée spécifiquement à la TAO, mais dont la TAO peut bénéficier, tout en s'appuyant éventuellement sur d'autres outils ou procédures ou algorithmes développés en commun ou hérités d'autres secteurs tels que la recherche documentaire, l'indexation automatique, la bibliométrie, le routage des messages, les systèmes experts, par exemple procédures pour isoler les paires, les triplets, compter les occurences et pondérer en conséquence, corréler pour déterminer ou analyser le contexte...

De tous ces logiciels ou linguiciels, l'utilisateur n'a bien entendu qu'une vue externe, celle du résultat et des délais, et son jugement, son acceptation ou son refus, vont s'exercer sur le résultat et la facilité qu'il aur à évaluer et à utiliser ce résultat, c'est-à-dire une traduction brute, plus ou moins rébarbative, qu'il aura à rendre intelligible. On voit ici toute l'importance que va revêtir l'ergonomie de la présentation de ces résultats et des conditions de travail qui seront offertes au traducteur ou à l'utilisateur final.

Selon les cas celui-ci sera appelé à réagir soit a postériori soit en ligne et en temps réel si le système est interactif. Cette interactivité ne peut se développer qu'en interne puisque ce n'est qu'au sein d'une même organisation que l'on peut veiller à créer des sécurités d'emploi pour éviter des entrées contradictoires de corrections ou d'additions, de bien paramétrer et de bien répondre aux questions posées par le système, ce qui nécessite une certaine spécialisation et une compétence linguistique.

Mais avec les gros systèmes serveurs de traduction assistée en ligne il n'est pas question non plus que chaque utilisateur puisse opérer librement en fonction de ses interêts propres, compte tenu du risque pour les autres utilisateurs. Navigation aids ne doit pas, du jour au lendemain,

donner SIDA...parce qu'un utilisateur du secteur biomédical a demandé une inclusion traitée sans précautions...

Pour l'utilisateur d'un système serveur la plus grande frustration vient aussi du temps qui s'écoule entre l'envoi de ses remarques et leur prise en compte.Une autre frustration vient de cette insuffisance actuelle d'analyse du contexte, qui laisse le système impuissant devant la polysémie, et l'insuffisance d'expressions courantes, qui devaient avoir été entrées comme des idiotismes avec leur équivalent dans l'autre langue, expressions que la machine est tout à fait incapable de reconstruire si l'expression n'a pas été reconnue et transposée en bloc (ex:subsidiairement et toutes choses égales d'ailleurs...)

La "post-édition"ou révision.

Ce terme barbare désigne la phase où le traducteur rend acceptable et intelligible un résultat où le bon se mêle au médiocre d'une façon telle qu'il est quelquefois difficile de démêler les fils. Autant il est aisé de repérer dans un original ou dans une traduction humaine une insuffisance de style ou de composition ou un contresens, autant il est délicat de mettre le doigt sur les insuffisances d'une traduction brute pour y localiser des défauts qui peuvent être de toute nature et parfaitement inattendus. La tâche étant particulièrement ingrate, il faut donc tout faire pour améliorer l'ergonomie de l'opération, par exemple faire apparaître en surbrillance les passages où la machine a hésité, ou a mis n'importe quoi, à tout hasard...ou encore avoir côte à côte le texte source et le texte cible correspondant, avec des aides au repérage de chaque phrase et un système de multifenêtrage pour afficher des données fournies par le dictionnaire, et pour pouvoir consulter des passages de traductions antérieures. Pour Systran, une excellente analyse des conditions de travail et de l'environnement de l'utilisateur est donnée par Pigott (17) qui montre toute l'importance d'un certain nombre de détails qui conditionnent le succès ou l'échec.

L'utilisateur.

Pour mieux comprendre et analyser les raisons du succès que la TAO commence à rencontrer, il faut en arriver enfin à parler du traducteur ou de l'utilisateur final puisque l'environnement commence et finit par l'utilisateur. Nous avons vu plus haut que celui-ci avait été maladroitement tenu à l'écart des développements de la TAO, alors qu'il aurait pu apporter beaucoup s'il avait été étroitement associé. Il faut dire que de son côté le traducteur a du mal à se faire entendre, dans une profession mal structurée, mal représentée.

Une enquête conduite en 1986 (18) révèle qu'on trouve en effet une grande majorité de traducteurs indépendants (free-lance), travaillant seuls, comme des artisans.

Ils ont indiqué le temps qu'ils consacraient à chaque type de tâche:

-10 à 15 % à des recherches de terminologie

-10 % à la préparation du texte ou "pré-édition"

-50 % environ à la traduction proprement dite

-15 % à la "post-édition"et à la mise en forme en vue de l'impression.

La même enquête montre que la plupart des textes sont techniques ou commerciaux.

Le texte arrive sous forme dactylographiée dans 54 % des cas, imprimée dans 33 % des cas, manuscrit (7 %), sur disquette ou bande magnétique ou par téléchargement (4

%), sur support audio(0,5 %).

Il est évident qu'avec la généralisation du traitement de texte et de l'édition électronique le texte sera dans bien des cas déjà numérisé et "propre", grâce aux correcteurs orthographiques ou autres outils disponibles en amont, de sorte que les pré-traitements vont se trouver fortement allégés et leur coût réduit. Il est clair aussi que la généralisation du recours aux banques de terminologie va aussi pénétrer l'environnement du traducteur, d'où réductions à prévoir sur les 10 % correspondants.

Il apparaît que, sans la TAO, c'est 50 % du temps du traducteur qui est passé à la traduction proprement dite, et 15 % dans la phase qui suit. L'intervention de la TAO? même si elle alourdit la "post-édition" en augmentant les 15 % réduit très sensiblement les 50 %. L'évolution sera relativement rapide puisque l'enquête révèle que plus de 50 % des traducteurs utilisent déjà en 1986 un microordinateur personnel pour le traitement de texte. Dans ces conditions on voit mal comment certains pourront reculer devant l'adjonction d'un bon logiciel de TAO.

Conclusion:

Après ce parcours un peu sinueux dans l'environnement technique de la TAO, quels sont les point essentiels qui méritent d'être regroupés et retenus? essayons de les énumérer:

La TAO n'est qu'une composante de l'industrie de la langue. Elle n'est plus une activité isolée mais peut bénéficier de recherches et de développements d'autres segments de cette industrie.

Le marché est important même s'il est difficile à évaluer. Il existe divers types de besoins et de marchés: documentation technique, serveurs d'information en ligne, applications liées à la synthèse et à la reconnaissance de la parole et à l'interprétation simultanée. Ici comme pour la traduction il s'agit de ne pas oublier de s'assurer le concours des interprètes, dont les mécanismes mentaux, en simultanée et en consécutive, qui isolent les idées, les concepts, au-delà des mots, en interdisant pratiquement le contresens, peuvent utilement être analysés.

Chaque nation, ou chaque groupe de nations d'une même langue, est amenée à "industrialiser" sa langue pour pouvoir conserver sa culture et son rayonnement.

Il ne faut pas attendre de la recherche un produit miracle qui dispenserait de tout effort de développement de la part des utilisateurs. La solution est certainement dans un regroupement de ces utilisateurs pour que ce développement soit effectué en commun, et donc à moindres frais.

Il faut intégrer au mieux la traduction dans la chaîne de traitement documentaire et favoriser la numérisation à la source en évitant le retour au support papier lorsque ce support n'est pas indispensable.

Il faut faciliter la tâche de l'utilisateur final et notamment du traducteur en veillant à une bonne ergonomie aux niveaux ou celui-ci intervient. Il faut aussi savoir et faire savoir que la TAO ne conduit pas à la disparition des traducteurs mais qu'elle modifie leurs conditions de travail dans le sens d'une aide de plus en plus précieuse, obtenue à leur profit etgrâce à eux, sous réserve que soit bien étudiée l'ergonomie des nouveaux postes de travail.

Il faut permettre une prise en compte rapide et aisée des remarques de l'utilisateur pour qu'il accepte de participer à l'évolution du système, en constatant des améliorations.

Dans le contexte dela TAO il faut favoriser aussi toute formule qui contribue à la réduction des barrières linguistiques, même lorsqu'il ne s'agit pas de TAO stricto sensu, et même si les objectifs sont très modestes: consultation en ligne et mise à jour fréquente d'outils linguistiques tels que dictionnaires classiques ou thésaurus multilingues, systèmes d'extraction de concepts, d'indexation et d'analyse automatique, ou de balayage rapide des textes, ou de routage, ou d'interrogation de banques de données à partir d'une autre langue..d'autant plus que les progrès réalisés dans ces domaines rejailliront sur la TAO, qui en bénéficie déjà largement. La base WTI (World Transindex) qui rassemble des signalements de traductions effectuées dans le monde (300000 références depuis 1977) et rend ainsi les traductions beaucoup plus accessibles mérite ici une mention spéciale.

Il faut limiter le secteur couvert par le système de TAO, et s'il n'est pas possible de se limiter à un secteur, développer les indicateurs de contexte qui font encore défaut, et ceci grâce au recours à l'intelligence artificielle.

Il faut agir en tenant compte de l'existence de plusieurs sortes d'utilisation, donc de plusieurs marchés de la TAO, distincts et nécessitant des environnements propres, et acceptés par les utilisateurs: par exemple:

-systèmes serveurs de traduction en ligne

-systèmes intégrés à une chaîne de traitement documentaire à l'intérieur de l'entreprise

-systèmes à l'usage individuel du traducteur indépendant..etc.

et enfin il faut ne pas se montrer trop optimiste ni trop sceptique, et savoir qu'il faudra en cette matière un effort soutenu et un investissement en relation avec la nature de l'enjeu.

oooOooo

Bibliographie:

1. M. Nagao. La traduction automatique. La Recherche. N° 150. Décembre 1983.

2. Chen Zhaoxiong et Gao Quinshi. English-chinese machine translation system IMT/EC. Coling 88. Budapest.

3.J. C. Guittet. commission des communautés européennes. La Jaune et la Rouge. Février 1989.

4. L. Danlos. La traduction automatique. Ann. Télécomm. 44. N° 1-2, 1989.

5. G. Dosi. Technical changes and survival: Europe semiconductor industry. Sussex European papers. N° 9. 1980.

6. Cahiers d'économie mondiale. Gerdic. Université de Rennes 1. Vol. 3, N° 1, Janv. Juin 1989.

7. C. Boitet. Representation and computation of units of translation for machine interprétation of spoken texts. ATR Interpreting Telephony Research Lab. Osaka. Sciences et Défense. Mai 1988.

8. K. N. Molholm. The DOD computer acquisition and logistic support (CALS) initiative. AGARD:TIP specialists' meeting. Brussels 1989.

9. Masaru Tomita et J. G. Carbonnel. The universal parser architecture for knowledge-based machine translation. Carnegie Mellon University (IJCAI 87).

10. Décision informatique. N° 138. Avril 1987.

12. Les industries de la langue. Editions DAICADIF. Septembre 1987.

13. Timbal-Duclaux. Revue communication et langage. Juillet 1989.

14. E. Launet. Les artisans de l'informatique. Sciences et technologies. N° 1. Janvier 1988.

15. O. Lavroff. Avis d'un utilisateur du système Systran. Encrages. N° 17.

16. Y. Zaharin. Toward an analyser (parser) in a machine translation system based on ideas from expert systems. GETA

presentation at the school of mathematics. Universitii Stains. Malaysia. Penang. Computational linguistics; Vol. 4. N° 2.

17. D. Smith. Translation practice in Europe. DEC. Translation and the computer. 8: a profession on the move. Conference proceedings. 1986.

18. G. Poirot. Le Japon affiche sa TAO, l'Europe cherche la sienne. C. R. MT Summit II de Munich. 01 Informatique. 4 Sept. 89.

19. Machine translation and machine-aided translation. W. J. Hutchins. Journal of documentation, Vol 34, N°2, June 1978.

20. Japon Panorama. Lettre de l'INIST. ISSN 0766-6195.

Figure 1. Histoire de la TAO: reperes.

```
     1940      1950      1960      1970      1980      1990
--------------------------------------------------------------------
         1946             1961     1970       1981         1995
       W. Weaver          CETA    Systran    Titus        Eurotra
       A. O. Booth                 Ohio   1977
                           1964           TAUM            JEIOA??
                    1957   ALPAC     1972   LOGOS
                   Sputnik         Winograd  ALPS   Bravice
                                   Fillmore  Wiedner

                                      1975    1982
                                     Systran  Gachot
                                   Luxembourg

                                       1978
                                       Ariane
                                       GETA
```

Figure 2. TAUM METEO: une simplicité biblique.

forecasts for ontario issued by environment canada at 11.30 am
est wednesday march 31st 1976 for today and thursday.

metro toronto
windsor.
cloudy with a chance of showers today and thursday.
low tonight 4. high thursday 10.
outlook for friday...sunny.
end.

prévisions pour l'ontario émises par environnement canada à 11 h
30 hne mercredi le 31 mars 1976 pour aujourd'hui et jeudi.

toronto et banlieue
windsor.
nuageux avec possibilité d'averses aujourd'hui et jeudi.
minimum ce soir 4.
high thursday 10.
aperçu pour vendredi...ensoleillé.
fin.

Figure 3. Industrie de la langue.

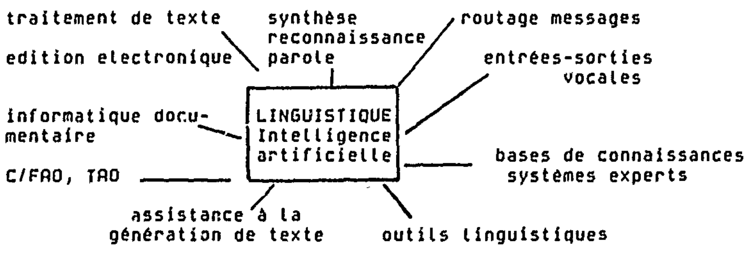

traitement de texte
edition electronique
synthèse
reconnaissance
parole
routage messages
entrées-sorties
vocales
informatique documentaire
LINGUISTIQUE
Intelligence
artificielle
C/FAO, TAO
bases de connaissances
systèmes experts
assistance à la
génération de texte
outils linguistiques

Figure 4: Occurences décroissantes
dans le fichier Inspec (1989)
Commande ..memt de Questel Plus

traitement de la langue	traduction
COMPUTATIONAL LINGUISTICS	LANGUAGE TRANSLATION
NATURAL LANGUAGES	NATURAL LANGUAGES
GRAMMARS	COMPUTATIONAL LINGUISTICS
LANGUAGE TRANSLATION	GRAMMARS
KNOWLEDGE REPRESENTATION	INDEXING
LINGUISTICS	LINGUISTICS
USER INTERFACES	CHEMICAL STRUCTURE
KNOWLEDGE BASED SYSTEMS	NOMENCLATURE
COMPUTER AIDED INSTRUCTION	ORGANIC COMPOUNDS
FORMAL LANGUAGES	CONTEXT FREE GRAMMARS
DATA STRUCTURES	GLOSSARIES
HIGH LEVEL LANGUAGES	KNOWLEDGE BASED SYSTEMS
KNOWLEDGE ENGINEERING	KNOWLEDGE ENGINEERING
LOGIC PROGRAMMING	MICROCOMPUTER APPLICATIONS
SPEECH RECOGNITION	SOFTWARE PACKAGES
ARTIFICIAL INTELLIGENCE	SPEECH RECOGNITION
COMPLETE COMPUTER PROGRAMS	CIRCUIT CAD
COMPUTER AIDED ANALYSIS	COMPUTER AIDED INSTRUCTION
COMPUTER SCIENCE EDUCATION	COMPUTERISED PICTURE PROCESSING
COMPUTERS	DIGITAL SIMULATION
CONTEXT FREE GRAMMARS	EXPERT SYSTEMS
DIGITAL SIMULATION	FORMAL LANGUAGES
DIRECTED GRAPHS	FORMAL SPECIFICATION
EDUCATIONAL COURSES	HYPERMEDIA
ERROR HANDLING	INFORMATION RETRIEVAL SYSTEMS
EXPERT SYSTEMS	INTERACTIVE SYSTEMS
EXPLANATION	LARGE SCALE INTEGRATION
FORMAL LOGIC	LEARNING SYSTEMS
FORMAL SPECIFICATION	
HEURISTIC PROGRAMMING	

Figure 5: Extrait du fichier PASCAL
de l'INIST. Exemple d'indexation multilingue

```
4/447 - (C) CNRS
NO   : PASCAL-INFODOC 89-0139407
ET   : A statistical approach to french/english translation
AU   : BROWN P; COCKE J; DELLA PIETRA S; DELLA PIETRA V; JELINEK F; MERCER R;
       ROOSSIN P
AF   : IBM res. div. T. J. Watson res. cent./Yorktown Heights NY 10598/USA
DT   : Congres; LA
SO   : RIAO 88: (Recherche d'Information Assistee par Ordinateur).
       Conference/1988-03-21/Cambridge MA; USA; DA. 1988; VOL. 2; PP.
       810-828; BIBL. 14 ref.
LA   : ENG
FA   : Approche de la traduction automatique qui utilise les techniques
       d'extraction d'information statistique dans de grandes bases de
       donnees.  Les correspondances entre termes sont baties a partir de la
       comparaison statistique d'un corpus et de sa traduction.  Les actes
       bilingues (francais-anglais) des sessions du parlement canadien ont
       servi de reference a cette etude
CC   : 205A04E
FD   : Traduction automatique; Methodologie; Analyse statistique; Francais;
       Anglais
ED   : Mechanical translation; Methodology; Statistical analysis; French;
       English
SD   : Traduccion automatica; Metodologia; Analisis estadistico; Frances;
       Ingles
```

Figure 6: TITUS.

Titus repose sur une méthode particulière de traduction dite à syntaxe contrôlée, qui n'autorise que l'emploi de formes d'expression obéissant à certains critères linguistiques restreints et prédéterminés. Les règles syntaxiques acceptées par Titus sont tout à fait naturelles et des plus classiques dans chaque langue. Les phrases ne doivent contenir que des termes figurant dans un dictionnaire préalablement établi. Chaque phrase est testée sur sa validité syntaxique et lexicale. Toute erreur ou toute ambiguité est signalée par l'affichaged'un message sur l'écran du terminal utilisé. En cas de polysémie, l'opérateur choisit la signification qui convient. Titus a été spécialement conçu pour le traitement multilingue des bases de données scientifiques et techniques.

SPECIMEN DE TRADUCTION BRUTE

COMMISSION
DES COMMUNAUTÉS
EUROPÉENNES

Direction Générale
Télécommunications Industries de l'Information et Innovation

ORIGINAL	TRADUCTION BRUTE SYSTRAN
Contribution for DG XIII brochure EUROTRA	Contribution pour la brochure de DG XIII EUROTRA
EUROTRA is a Community research and development programme for the creation of a machine translation system of advanced design capable of dealing with all the official languages of the EC. It was adopted by Council Decision 82/752/EEC of 4 November 1982 and extended by Council Decision 86/591/EEC of 26 November 1986 to include Spanish and Portuguese following the accession of Spain and Portugal.	Eurotra est un programme communautaire de recherches et de développement pour la création d'un système de traduction automatique de conception avancée capable de traiter de toutes les langues officielles de la CE. Il a été adopté par la décision 82/752/EEC du Conseil du 4 novembre 1982 et élargi par la décision 86/591/EEC du Conseil du 26 novembre 1986 pour comprendre espagnol et portugais après l'adhésion de l'Espagne et du Portugal.
The programme is jointly financed by the Community and its member States. Its objective is the creation of a prototype system which would be operational for a limited subject field and for a limited number of text types with a vocabulary of approximately 20.000 entries. This will provide the basis for development on an industrial scale in the period following the current programme. In addition EUROTRA aims at creating in Europe a "critical mass" of expertise in machine translation and computational linguistics in general.	Le programme est conjointement financé par la Communauté et ses Etats membres. Son objectif est la création d'un système prototype qui serait opérationnel pour un domaine limité et pour un nombre limité de types de texte avec un vocabulaire d'approximativement 20,000 entrées. Ceci fournira la base pour le développement sur une échelle industrielle pendant la période après le programme actuel. En plus EUROTRA vise à créer en Europe une "masse critique" des connaissances dans la traduction automatique et de linguistique computationnelle en général.
EUROTRA is a seven-year programme divided into three phases, each with its own tasks and objectives:	Eurotra est un programme de sept ans divisé en trois phases, chacun avec ses propres tâches et objectifs:
A. The preparatory phase (two years) during which:	A. La phase préparatoire (deux ans) pendant laquelle:
(1) the organizational arrangements for the project were agreed,	(1) les dispositions organisationnelles pour le projet ont été convenues,
(2) the linguistic and software specifications were defined.	(2) linguistiques et les spécifications de logiciel ont été définies.

> " Humans have not only the ability to parse sentences but also the ability to recognise strings as ill-formed."
>
> Flickinger, Nerbonne, Sag & Wasow

Essai de typologie des textes source, dans le cadre de la traduction assistée par ordinateur

par

Jean Gordon

Directrice des Services centraux
Direction générale des opérations de traduction
Langues officielles et Traduction
Secrétariat d'Etat du Canada
Ottawa (Ottawa)
K1A 0M5

SOMMAIRE

Le recours à la traduction assistée par ordinateur est une décision réfléchie. Tous les facteurs doivent être pris en considération. Une typologie des textes permet de déterminer la catégorie à laquelle ils appartiennent et leur degré de "taoisabilité". Elle se fonde sur les caractéristiques matérielles (le support utilisé et sa compatibilité), les caractéristique terminologiques (la variété, la complexité, et la stabilité du vocabulaire) et les caractéristiques stylistiques (la présence de certains traits linguistiques qui affectent la qualité des résultats). Cet examen du profil des documents sert d'outil de sélection et de tri et il doit être assorti d'une analyse approfondie avant la décision finale.

SUMMARY

The question of whether or not to use computer-assisted translation requires serious thought. All factors must be taken into consideration. A typology helps to identify the category into which texts fall and the extent to which they lend themselves to this approach. The typology is based on material considerations (medium used and compatibility), terminological considerations (variety, complexity ans stability of vocabulary) and st .istic considerations (presence of certain linguistic characteristics that affect the quality of results). This examination of the profile of documents is a sort of pre-screening tool. A thorough analysis should also be conducted before any final decision is taken.

Introduction

Tout texte peut être traduit. Pour un service de traduction, qui ne rejette pas les documents pour cause de rhétorique, d'emphase ou d'orthographe, tout texte doit être traduit.

Mais si le traducteur humain peut affronter, fut-ce au prix d'une perte de productivité, le traitement de documents aux caractéristiques les plus diverses, la machine en est incapable.

Il est admis que le traducteur professionnel doit satisfaire aux exigences des organismes ou des clients. Des outils ont été mis au point qui permettent de sélectionner celui qui saura le mieux répondre à ces besoins.

La situation est toute différente en ce qui concerne l'ordinateur et les logiciels de traduction assistée. D'abord, aucun fabricant sérieux ne prétend que son système peut traduire indifféremment tout ce qui passe. Ensuite, la plupart des organisations n'ont pas d'outils ou d'examens type à faire passer aux systèmes commercialisés actuels, ou à ceux que nous promet l'avenir. Et avant même de choisir un système, il faut d'abord déterminer si les documents à traiter se prêtent à ce genre de traduction. En effet, l'introduction d'un système impose des contraintes à l'organisation. Si elle ne bouleverse pas toujours l'organisation et les méthodes de travail, elle oblige au moins à des aménagements. Elle suppose aussi un investissement plus ou moins important.

L'évaluation du caractère taoisable des textes visés prend donc toute son importance lors de la décision initiale, lorsqu'il faut répondre à la question: La tao convient-elle aux textes, ou encore, les textes se prêtent-ils à la tao?

Le présent essai de typologie des documents aux fins de la tao n'a pour objet que de faciliter leur tri, une première sélection. S'il est utile, ce serait au même titre que les outils de pré-sélection des candidats qui se présentent à un poste de traducteurs.

Ceux qui répondent au profil établi sont admis à l'examen. De même, cette typologie sommaire vise à faciliter l'établissement du profil des textes dans l'optique de leur taoisation éventuelle, avant une évaluation plus approfondie assortie d'essais.

Les facteurs décrits ci-dessous ont été retenus en raison de leur influence sur les résultats de la traduction automatique ou de la traduction assistée par ordinateur, à partir d'un texte source anglais et pour un texte cible français. Ces résultats ont été constatés au cours d'essais de plusieurs mois avec trois systèmes commerciaux, à savoir ALPS et Systran (CEE), au Secrétariat général de l'OTAN à Bruxelles, et LOGOS au Secrétariat d'Etat du Canada. Ils sont confirmés par des analyses effectuées par ailleurs ou documentées dans les ouvrages mentionnés en référence.

Le volume à traduire n'a pas été pris en considération. Toutefois, ce critère doit peser très lourd dans la balance lorsqu'on analyse la rentabilité de la tao, ou de sa rentabilisation éventuelle.

Trois types de facteurs influent sur la traduction assistée par ordinateur et sa qualité, à savoir l'aspect matériel ou technique, la terminologie et les caractéristiques stylistiques.

Facteurs matériels ou techniques

Il y a peu d'années, les spécialistes tenaient que tous les documents à taoiser devaient être disponibles sur support magnétique.

C'est une condition essentielle, puisque l'ordinateur ne peut lire que la forme ordinolingue. Ce n'est pas une condition suffisante, puisque l'ordinateur ne peut pas tout lire, si la conversion est mauvaise ou impossible. A l'heure de la multiplicité des postes de travail, et des systèmes, il convient d'être plus spécifique. Il ne suffit plus que le texte source ait été dactylographié sur un ordinateur ou avec un traitement de textes quelconque, encore faut-il que le fichier soit convertible et que les codes de formatage ou autres le soient aussi.

Depuis, l'apparition de lecteurs optiques très performants, avec capacité d'apprentissage, permet une transition éventuelle entre systèmes, ou encore la taoisation de textes pour lesquels des disquettes ne sont pas disponibles ou ne le sont plus.

Ce facteur technique, souvent sous-estimé, a un impact considérable sur les opérations.

Il faut souligner en particulier les problèmes que peuvent causer lors de la conversion les codes relatifs aux tableaux, aux graphiques et même à la disposition en colonnes. Si une solution technique éprouvée n'est pas disponible, cette caractéristique doit être prise en compte lors de l'analyse linguistique des documents.

Une conversion imparfaite ou des problèmes constants de formatage diminuent la productivité et rallongent les délais de traduction et de livraison des textes. En outre, même si la compatibilité existe en théorie, des règles précises de dactylographie et de formatage doivent être repectées. Ces règles diffèrent fort peu des règles de traitement de textes, cependant l'expérience a révélé que celles-ci sont souvent oubliées par le personnel de soutien, si seule la sortie papier est importante. Pour pallier à cet oubli, des séances d'information ont été mises au point au Secrétariat d'Etat du Canada pour les personnes appelées à dactylographier les textes source des projets de tao.

Si les conditions matérielles ne sont pas bonnes, il est préférable de renoncer à introduire la tao, ou du moins de reporter son implantation. Sauf dans des circonstances exceptionelles, ce facteur est donc éliminatoire.

Facteurs terminologiques

Dans le contexte de la tao, les facteurs terminologiques ont un impact considérable.

En effet, les systèmes sont généralement livrés avec un dictionnaire de base. Mais comme tout traducteur qui se respecte ne peut se limiter au contenu d'un seul dictionnaire général, il ne saurait utiliser la tao avec le dictionnaire standard. Le dictionnaire du système doit donc être enrichi de la terminologie propre au domaine et à l'institution.

Si cette opération d'entrée dans le dictionnaire est parfois très simple et très rapide, elle peut aussi exiger une codification complexe ou de multiples entrées. Un terme peut prendre quelques minutes, ou plusieurs heures s'il faut en valider le codage.

Dans un premier temps, il s'agira sans doute de mots ne figurant pas du tout au dictionnaire, signalés par une recherche de mots nouveaux dans un corpus plus ou moins étendu.

Cette recherche est faite avant la traduction. Sauf pour les verbes dont le codage est fort complexe et doit généralement être confié aux spécialistes, l'entrée de mots nouveaux est facile. En outre, même dans un domaine qui n'a pas déjà été traité par la machine, leur proportion est relativement faible. A titre d'exemple, pour une page de 595 mots choisie au hasard dans *Totem poles according to crests and topics* (Marius Barbeau, Secrétariat d'Etat, Ottawa 1964), Logos a recensé vingt-trois mots non trouvés au dictionnaire, dont sept sont des noms communs:

Si'aks	001	
stench	001	
tsawltsap	001	
volcano	002	
Weehawn	001	
Wigyidemrhsaek	001	(001E)

Dans un deuxième temps doivent être identifiés les mots utilisés dans une acceptation différente de celle du dictionnaire standard mais propre au domaine ou à l'organisme visé. Cette étape est postérieure à un premier passage machine qui permet de repérer les écarts. Dans deux phrases du passage cité ci-dessus, "traducteur" et "spot" ont un équivalent tiré du domaine informatique et devraient être remplacés dans le domaine pertinent:

William Beynon acting as interpreter.
Guillaume? Beynon qui fait fonction du traducteur. (002E)
The guide pointed to a spot nearby.
Le guide a indiqué un spot tout près. (003E)

En troisième lieu, il faut affronter le phénomène de la polysémie intrasectorielle, qui augmente avec la complexité du domaine. En effet, il y a une limite aux distinctions entrées dans le système et on arrive vite au point de saturation. Aussi, de tous les problèmes terminologiques, celui de la polysémie est l'un des plus complexes, et les divers logiciels n'ont pas encore trouvé d'arme absolue.

Cancel the check.
Annuler le chèque.
Annuler la vérification. (004E)

L'investissement terminologique, équivalant au travail du traducteur qui dans un autre contexte consigne ses recherches sur fiche, peut exiger beaucoup de temps et d'efforts dont les exemples très simples fournis ci-dessus ne donnent qu'une idée partielle. Ce travail sera plus ou moins long suivant l'étendue du vocabulaire utilisé, elle-même fonction du type de textes et de l'ampleur du domaine. Ainsi, la terminologie d'un corpus composé de comptes rendus de réunions portant sur le transport des explosifs sera bien plus limitée que celle d'un corpus mixte de comptes rendus, d'articles et de rapports sur le même sujet, elle même dépassée par l'ampleur d'un corpus semblable sur la mécanique.

Si l'on se fonde uniquement sur la délimitation du domaine pour déterminer la variété probable de la terminologie du corpus dont la taoisation est envisagée, il convient d'en évaluer également l'uniformité du vocabulaire. Pour le même domaine, l'intervention d'auteurs différents peut entraîner une instabilité de la terminologie ou introduire des variantes qui n'ont pas d'impact sur la compréhension du texte par le lecteur humain, mais affectent la traduction machine. Or, la stabilité ou la cohérence de la terminologie rentabilise plus rapidement le temps consacré à chaque entrée au dictionnaire et la qualité de la traduction subséquente s'en trouve améliorée.

Un succès de la traduction automatique souvent cité, c'est-à-dire le système METEO au Canada, est notamment attribuable à la terminologie limitée du sous-langage des bulletins météorologiques.

Imposer un langage contrôlé est une façon d'obtenir la stabilité recherchée et certains organismes ont eu recours à cette méthode. Dans certains contextes organisationnels, elle est cependant impossible à appliquer. Avec la généralisation des postes de travail chez les cadres, l'utilisation d'un dictionnaire commun serait un compromis à envisager.

Plusieurs méthodes permettent d'évaluer le facteur terminologique. De façon générale, on peut présumer que plus un domaine est circonscrit, plus la terminologie est limitée et le sous-langage clairement défini. On peut procéder de façon empirique et analyser les textes disponibles, ou se fier à sa connaissance approfondie de la demande.

Pour compléter une évaluation, des outils sont disponibles qui mesurent la variété terminologique d'un corpus. Ils s'apparentent aux listes de mots en contexte utilisés par les juristes (KWIC). Certains fabricants de tao ont intégré ces logiciels et offrent des listes de fréquence, avec ou sans contexte.

Facteurs stylistiques

La dernière analyse, celle du style, porte sur une caractéristique beaucoup plus difficile à cerner et à définir.

Les typologies d'évaluation des systèmes de traduction, et les nombreux essais et exemples disponibles, donnent un tableau assez complet, mais sans doute pas exhaustif, des embuches et des problèmes que posent encore les textes source.

Il est communément admis que certaines catégories de documents, au style très personnalisé, mais aussi au vocabulaire difficilement limitable, ne se prêtent pas à la tao. Aussi, les caractéristiques retenues et décrites ci-dessous ne sont nullement pertinentes dans le cas d'ouvrages littéraires ou de textes juridiques où, si les erreurs sont en principe rares, le style est souvent recherché, sinon dense.

En ce qui concerne les textes techniques, scientifiques ou administratifs, et plus particulièrement les textes de type informatif, leurs caractéristiques peuvent être cernées de façon générale et l'on peut parler de textes bien structurés ou bien écrits, idiomatiques ou rigoureux.

L'analyse linguistique décrite ici, de portée fort modeste, apporte une dimension linguistique, essentielle à une typologie d'évaluation de corpus aux fins de la tao. Conçue de façon pragmatique, elle peut porter sur un volume variable de textes disponibles; sa fiabilité est toutefois directement fonction non pas tant du volume de textes analysés mais de leur représentativité.

Aux fins du présent essai de typologie, seules quelques caractéristiques linguistiques ont été retenues. D'autres pourraient être ajoutées et le modèle raffiné. Celles énumérées ici sont celles qui ont un impact immédiat sur la qualité et présentent l'intérêt d'être faciles à reconnaître. Elles peuvent être utilisées par une personne étrangère à la tao, et un traducteur habitué à un système d'évaluation de la traduction n'a aucun mal à les appliquer.

Ces caractéristiques ont toutes un trait commun : leur présence a un effet négatif, soit qu'elle pose des problèmes au niveau du dictionnaire par la création d'homographes, soit qu'elle fausse l'analyse ou la complique. Les unes sont considérées par les grammairiens et les linguistes comme des fautes. Les autres sont des particularités stylistiques parfaitement acceptables, ou même recherchées.

Certains pourraient s'étonner de voir y figurer en bonne place des erreurs d'orthographe, de grammaire ou d'usage. Cette inclusion reconnaît la réalité du monde imparfait des rédacteurs pressés, ou qui n'écrivent pas dans leur langue. Les exemples donnés sont tirés de textes réels.

Les fautes d'orthographe, auxquelles peuvent être assimilées les fautes de frappe et coquilles, créent des mots inconnus, non traduisibles:

... metal, cables, elements ect.
du métal, des câbles, des éléments ?ect. (303)
He has responsability ...
Il a le? responsability ... (L810)

Pis encore, en remplaçant un mot par un autre, elles font déraper l'analyse en déguisant un verbe en un nom, ou un article en un verbe:

One of the functions is each category is designated ...
Une des fonctions est est désignée chaque catégorie...(71)

Les fautes d'accord les plus graves, aux fins de la taoisation, sont celles du sujet et du verbe; plus fréquentes que l'on n'imagine, elles faussent généralement l'analyse si elles créent une ambiguïté quant au sujet du verbe:

This group of students wish to visit the museum.
Ce groupe du souhait d'élèves pour visiter le musée. (005E)
If either A or B wish to go out, we will do this
Si souhait A ou B pour sortir, nous ferons ceci. (006E)

Les erreurs que des réviseurs ou des évaluateurs linguistiques eussent sanctionnées ne figurent pas, rappelons-le, parmi les fautes à relever, si elles n'ont pas d'effet ou ont au contraire un effet bénéfique en clarifiant des liens grammaticaux. C'est le cas de certains gallicismes ou encore de compléments de nom.

Dans une toute autre catégorie tombent les ellipses et omissions. Fautives ou justifiées, elles ont un effet néfaste incontestable.

L'omission de l'article devant un nom peut créer un homographe stylistique, et en l'absence de l'information voulue, le logiciel traduira un nom par un verbe, ou inversement:

Paint surface between lines.
Surface de peinture entre les lignes. (007E)
Fill vase with water.
Vase de plein avec l'eau. (008E)

L'omission du pronom sujet devant un verbe peut également créer un homographe stylistique, avec le même genre de conséquence que ci-dessus:

To produce reports,
Le produit rapporte, (267P)

Les autres types d'omissions sont moins répandues. Elles peuvent être relevées quand même, mais leur effet semble plus aléatoire.

If, however, a function other than the prime desired.
Si, toutefois, une fonction autre que l'apogée voulue.
(73)

Close the back and the front door.
Fermer le dos et la porte de devant. (009E)
How to turn on and off the motor.
Comment tourner le moteur de marche/arrêt. (010E)

L'utilisation des prépositions est à relever. La plupart des logiciels ont intégré les règles ou les constructions idiomatiques les plus courantes. Cependant, la variété des formes adverbiales consistant d'un verbe et d'une préposition est infinie et chaque auteur peut en ajouter à sa guise. Les règles nécessaires se contredisent alors, ou créent des unités lexicales contradictoires, avec un résultat à l'avenant:

The sytem manager can go on to delete several entries
Le gestionnaire du système peut être supprimé continuer plusieurs entrées
(267P)

Une caractéristique particulièrement meurtrière est l'empilage de compléments ou d'attributs, de noms ou d'adjectifs, dont la présence semble inciter les systèmes à des distributions qu'on pourrait penser aléatoires:

The site level troubleshooting tasks are essentially performed...
L'installation nivelle est essentiellement exécuté les tâches de dépannage....
(258)
(The troubleshooting tasks at site level are essentially performed / Les tâches de dépannage, au niveau d'installation, sont essentiellement exécutées)
"Appendix A" "Maintenance Technical Parameters Check Sheet" should be used.
La "feuille de vérification de paramètres de maintenance d'"A" d'annexe "technique" devrait être utilisée. (316)

La présence de propositions relatives en cascade a aussi été incluse, même si dans certains textes bien structurés elle ne semble pas nuire au résultat. Ce n'est toutefois pas la règle mais l'exception.

L'utilisation d'abréviations et d'acronymes, et plus particulièrement d'acronymes homographes, forme une catégorie distincte. Ces derniers, plus faciles à prononcer, sont très populaires. La question des acronymes pourrait être considérée parmi les facteurs terminologiques. Elle figure parmi les facteurs stylistiques en raison des difficultés particulières que peut présenter le codage ou l'entrée au dictionnaire de ces termes, mais aussi parce qu'elle donne lieu à des erreurs ou des irrégularités d'usage ou de graphie.

Shape delegates arrived.
Les déléguées de forme sont arrivées. (012E)
The SCC panels offer status information.
Le SCC lambrisse l'information d'état d'offre. (17)

Pour finir, la présence de tableaux ou graphiques, ou la présentation en colonnes, si elle n'a pas été prise en compte au chapitre des considérations matérielles et réglée à ce stade par des dispositions idoines, doit l'être au moment de l'analyse stylistique.

Les traits recensés ci-dessus peuvent être fréquents, occasionnels ou rares, et le corpus sera coté en conséquence. La cote peut être combinée, c'est-à-dire porter sur l'occurence des diverses caractéristiques. La marge de tolérance sera alors plus large. A titre indicatif, on suggère que la catégorie "rare" soit réservée à moins de trois occurences par page, la catégorie "occasionnelle" à trois à six occurences par page, et la catégorie "fréquente" à plus de six occurences par page. La cote peut aussi être attribuée à chacune des caractéristiques linguistiques, et la marge de tolérance sera resserrée en conséquence.

Cependant, il ne faut pas oublier que, quel qu'il soit, le système de traduction utilisé, il est raffiné par l'entrée de règles et de terminologie, et c'est là un effort tous les jours renouvelé.

L'analyse linguistique du système sera donc complétée par une évaluation de la stabilité du style et de la récurrence des formes typiques, fondée sur les documents mêmes, ou induite de la structure organisationnelle et des usages de l'institution. Ainsi, certains organismes utilisent des formules figées pour tous leurs documents, ou leurs procès-verbaux. Dés types de documents sont généralement rédigés par le même auteur, ou par le même groupe d'auteurs, atteignant une cohérence de style qu'on ne saurait attendre de rédacteurs éparpillés dans tout le pays, ou même dans plusieurs pays.

Dans une situation de ce genre, l'intégration sous une forme ou une autre des expressions récurrentes et des tournures particulièrement fréquentes est possible et peut donner des résultats intéressants.

On peut d'ailleurs envisager d'influer sur le style, en collaborant étroitement avec les auteurs, sans pour autant imposer un carcan de rédaction, et de modifier le profil des textes. Une fois prise la décision de se lancer dans la tao, une séance d'information atttire l'attention des rédacteurs sur certains types de problèmes qui auront été identifiés dans le corpus de base et seront évités par le respect de certaines règles de rédaction.

Grille d'évaluation

En appliquant ces trois facteurs, on crée une grille d'évaluation qui permet de classer les documents en sept catégories, par ordre décroissant de "taoisabilité".

Suivant les objectifs poursuivis, une fois reconnu le profil, une décision organisationnelle pourrait en principe faire passer un corpus à la catégorie supérieure, en imposant par exemple aux cadres un vocabulaire uniforme et un style contrôlé. Encore faudrait-il que cette révolution soit envisageable et ses effets assurés.

Textes très taoisables

Ce sont les textes éminemment taoisables, idéaux sur tous les plans et au regard des divers facteurs.

En l'occurence, il s'agirait de textes entrés directement sur l'ordinateur de la tao, avec tous les codes de formatage voulus. Au stade actuel, cela pourrait impliquer que le document consiste en texte continu et ne comporte aucune colonne, pas un seul tableau et évidemment pas de graphiques.

La terminologie en serait limitée et stable, sans polysémie, ce qui serait le cas pour un sous-domaine au sous-langage bien défini.

Sur le plan stylistique, les textes ne présenteraient aucune des caractéristiques négatives: texte sans ambiguïtés, rédigé vigoureusement suivant les règles grammaticales, respectant l'usage, sans fautes d'orthographe ni coquilles. Les phrases sont courtes, mais sans raccourcis, ni ellipses. Si les empilages et les cascades de mots en sont bannis, sont également exclues les constructions idiomatiques où les prépositions abondent. La fantaisie et l'imagination ne viennent pas perturber ce portrait idyllique.

Pour un corpus composé uniquement de textes de ce genre, satisfaisant au critère relatif à la terminologie et donc au sous-langage bien délimité, la difficulté pourrait plutôt être d'un autre ordre, à savoir l'existence d'un volume suffisant.

Textes généralement taoisables

Ce sont les textes qui répondent de façon générale à tous les critères. Certains écarts en viennent compliquer le traitement ou ralentir le processus, mais ils sont bien délimités et peuvent être corrigés par une intervention précise.

Par exemple, il pourrait s'agir de documents entrés dans le même ordinateur, à la terminologie très limitée, mais dont les caractéristiques linguistiques négatives sont extrêmement stables et peuvent être intégrées dans un système de tao. C'est le cas des bulletins météorologiques, dont le profil très particulier a donné lieu au développement d'un système ciblé, ce qu'autorise le volume annuel considérable.

Entreraient aussi dans cette catégorie des documents d'un domaine très limité, au profil linguistique positif (occurences rares), mais seulement disponibles sur support magnétique. La solution, soit l'établissement ou le perfectionnement de la conversion, serait alors d'ordre technique.

Textes taoisables

Ce sont les textes qui répondent à la plupart des critères énumérés. Les écarts qui affectent les résultats ne se prêtent pas une solution unique précise mais les efforts de correction doivent se poursuivre sur une certaine période ou peuvent porter sur plusieurs aspects.

Par exemple, il s'agirait de textes d'un domaine limité, et au profil stylistique à la cote rare, dont les caractéristiques négatives n'ont pas une fréquence significative. La solution serait d'ordre terminologique et linguistique.

Seraient taoisables aussi les textes sur support magnétique mais encombrés de colonnes, ou encore ceux sur papier, mais dont les caractères peuvent être lus par lecteur optique. La solution serait alors d'ordre technique et organisationnel.

Textes peut-être taoisables

Ce sont les textes qui répondent généralement à la plupart des critères énumérés. Des écarts affectent les résultats mais il pourraient être réglés par du travail portant sur l'aspect linguistique.

Il pourrait s'agir de textes dans un domaine à la terminologie limitée, aux occurences stylistiques occasionnelles mais récurrentes et uniformes. Par exemple, l'utilisation d'acronymes ou d'abréviations est fréquente, mais cohérente, ou encore le texte contient des constructions verbales particulières, qui paraissent stables et peuvent faire l'objet d'une règle. La solution est d'ordre linguistique mais peut être coûteuse.

C'est pour cette catégorie que l'analyse linguistitique plus poussée est essentielle, et elle doit être assortie d'une évaluation du coût des améliorations nécessaires.

Textes difficilement taoisables

Ce sont des textes qui ne répondent pas aux critères. Leurs caractéristiques matérielles, terminologiques ou stylistiques affectent le traitement.

Ainsi, serait difficilement taoisable un ensemble de manuels sur des sujets divers, en raison du vaste domaine et de la terminologie variée et instable.

Sur le plan matériel, des textes aux nombreux tableaux et graphiques présenteraient le même profil de difficulté.

Textes très difficilement taoisables

Ce sont les textes qui ne répondent pas à l'un des critères essentiels. Leurs caractéristiques terminologiques et stylistiques affectent considérablement le traitement.

Entrerait dans cette catégorie, par exemple, un corpus de jugements d'un tribunal administratif fiscal. Le domaine quoique précis est vaste, et les auteurs sont nombreux avec toute latitude pour exercer leurs prérogatives rédactionnelles. Le facteur stylistique serait éliminatoire.

Textes non taoisables

Ce sont les textes qui ne répondent pas aux critères essentiels. Leurs caractéristiques sont éliminatoires: Sur le plan terminologique, un domaine non défini, sur le plan stylistique, des caractéristiques négatives fréquentes et non cohérentes, et sur le plan matériel une forme non ordinolingue.

Pour des textes informatifs, ce serait le cas de publications sur les marques de commerce, dont la terminologie couvre pratiquement tous les domaines, sans aucune uniformité, et dont le style elliptique pose de considérables problèmes d'analyse.

Ce serait évidemment le cas aussi d'articles disponibles sur papier seulement, des manuscrits par exemple, non lisibles par lecteur optique.

Conclusion

L'essai de typologie présenté ici n'est qu'une ébauche, qui reste à raffiner et à préciser. L'obstacle considérable que présente la multiplicité des types de textes et de leurs caractéristiques n'a pas été franchi. D'autres analyses plus savantes ont traité du sujet. Dans un parti pris de simplification, le développement a été tenté d'un outil qui serve de tamis au gestionnaire à la croisée de la technologie.

Cet outil devra évoluer, d'abord pour s'adapter à l'évolution rapide du secteur informatique car l'on peut espérer que les problèmes de conversion seront réglés un jour. La typologie pourra s'enrichir aussi sur le plan linguistique à mesure qu'elle est utilisée sur des corpus différents. La quantification plus précise des occurences serait notamment précieuse pour situer les corpus. Des études en cours sur les évaluations donneront une information supplémentaire. D'autres articles sur le traitement des langues naturelles fourniront des données de comparaison.

Cet outil donc, imparfait mais perfectible, peut aider le gestionnaire à déterminer dans un premier temps si la traduction assistée par ordinateur est une solution envisageable.

Pour quelques cas clairs, elle lui évitera sans doute une analyse plus poussée. Dans d'autres situations, elle ne remplacera pas une analyse des structures linguistiques des documents, ou à défaut une évaluation d'essais contrôlés de tao, afin de cibler l'application de celle-ci.

Cette évaluation des facteurs intrinsèques doit être complétée, cela va de soi, par une analyse rigoureuse de tous les facteurs externes, dont le contexte et les impératifs organisationnels, le coût d'investissement et d'exploitation, et surtout les objectifs poursuivis.

La traduction automatique ou assistée par ordinateur offre une solution tentante aux problèmes que peut poser la gestion de la charge de travail en traduction. Avant de céder à la tentation, il est sage de mesurer ses chances de succès.

Bibliographie sommaire

Bourbeau Laurent, Kittredge Richard, 1988
Projet d'automatisation d'accès aux brevets et inventions : Traduction assistée par ordinateur, Rapport final présenté au Bureau de la propriété intellectuelle du ministère de la Consommation et des Corporations, Odyssée Recherches appliquées, Montréal, mars 1988

Bourbeau Laurent, 1990
Elaboration et mise au point d'une méthodologie d'évaluation linguistique de systèmes de traduction assistée par ordinateur, Montréal, mars 1990

Flickinger Daniel, Nerbonne John, Sag Yvan et Wasow Thomas, 1987
Toward Evaluation of NLP Systems, in Forum of the Association for Computational Linguistics - Evaluating Natural Language Processing Systems, Hewlett Packard Laboratories, Palo Alto, juin 1987

Gordon Jean, 1989
Sélection d'un site TAO, Facteurs et critères, Ottawa, septembre 1989

Gordon Jean, 1989
Variation du coût du mot en TAO, Exercice théorique, Ottawa, août 1989

Gordon Jean,1990
Grille et méthodologie d'évaluation de versions successives d'un système de traduction assistée par ordinateur, Ottawa, avril 1990

Kaji Hiroyuki, 1989
Language Control for Effective Utilization of HICATS/JE, Systems Development Laboratory, Hitachi Ltd., in MT Summit II, Munich, août 1989.

Lehrberger John, Bourbeau Laurent, 1988
Linguistic Characteristics of MT Systems and General Methodology of Evaluation, John Benjamins Publishing Company, Amsterdam, Philadelphie, 1988

PROBLEMS ENCOUNTERED WITH THE USE OF COMPUTER ASSISTED TRANSLATION IN A TECHNICAL PUBLICATIONS PRODUCTION ENVIRONMENT

by

D.J. Pinna
Translation Manager
Aircraft Division Product Support (M9131)
AEROSPATIALE
316, route de Bayonne BP 3153
Toulouse Cedex
31060
France

SUMMARY

The purpose of this paper is to give a brief overview of the problems evidenced during studies into the use of Computer Assisted Translation (CAT) in the AEROSPATIALE AIRCRAFT DIVISION technical publications production environment. The aim is not to review the capabilities of the various CAT systems available in a comparative study but rather to highlight the technical, economical and psychological problems that have to date precluded the integration of CAT in this very specific industrial context.

1. INTRODUCTION

A. GENERAL

In 1989, for the first time, AEROSPATIALE, with their European partners, logged more than 30% of world orders for civil aircraft in their sector of the market thus becoming the second largest aircraft manufacturer in their category.

In that year there was a total fleet of 721 AIRBUS and ATR aircraft in service with 142 customers.
By the year 1995, in just 5 years time, this fleet will have grown to around 2500 aircraft in regular service around the world with some 300 customers.
Such rapid expansion implies a constant search for increased productivity and efficiency to achieve reduced production costs and cycles while continuously enhancing the quality of service provided to the customers.

The AEROSPATIALE AIRCRAFT DIVISION, fully aware of the commercial stakes involved, is organizing to meet this challenge and has adopted a strategy largely based on :

- an ambitious training program aimed at adapting the personnel to the new requirements of the rapidly evolving industrial context.
- the development of advanced production means taking maximum advantage of the possibilities offered by computerization.

It is in this highly dynamic context characterized by a constant search for new means of pushing back the limits of productivity and efficiency that the Technical Publications Department has conducted numerous studies into the possible utilization of Machine Translation (MT) or Computer Assisted Translation (CAT).

B. HISTORY OF TRANSLATION IN THE AEROSPATIALE AIRCRAFT DIVISION

Translation in the Aircraft Division has undergone and is still undergoing considerable change to adapt to the requirements of a continuously evolving environment.

It was with the success of CARAVELLE in the fifties that the need for translation led to the setting up of small groups of translators within various departments : Design Office, Production, Quality Assurance, Flight Test and, of course, Product Support in direct contact with the customer.

These small groups grew in size with the first experience of European cooperation : CONCORDE.

At that time only a very small percentage of AEROSPATIALE personnel was capable of getting along in English. It was therefore necessary to translate all the documents transiting between BRITISH AEROSPACE and AEROSPATIALE. The presence of an interpreter was indispensable at all the working meetings between the two partners.

The CONCORDE technical publications were initially issued in the two official languages of the program, English and French, before the French version was abandoned for cost reasons.

When the AIRBUS program was launched, AEROSPATIALE, rich with the experience gained with CONCORDE, strove to eliminate the problems arising from the use of two (or more) languages.
At that time only a very small percentage of AEROSPATIALE personnel was capable of getting along in English. It was therefore necessary to translate all the documents transiting between BRITISH AEROSPACE and AEROSPATIALE. The presence of an interpreter was indispensable at all the working meetings between the two partners.

The CONCORDE technical publications were initially issued in the two official languages of the program, English and French, before the French version was abandoned for cost reasons.

When the AIRBUS program was launched, AEROSPATIALE, rich with the experience gained with CONCORDE, strove to eliminate the problems arising from the use of two (or more) languages.

English was adopted as the official language of the AIRBUS program. All correspondence between the partners had to be in the English language. AEROSPATIALE therefore launched a vast training program so that the personnel concerned acquired a level of English sufficient :

- to understand routine correspondence received in English,
- to write directly in English simple memos addressed to the partners,
- to get along in inter-partner meetings conducted in English without the assistance of an interpreter.

Thus, in many sectors, the use of English as the official language for the AIRBUS and, later on, the ATR programs resulted in a reduction in workload for the translators and the gradual dissolution of the translation offices.

The Product Support translation office, however, largely due to the volume of translation involved in the production of technical publications continued to grow and has now become the largest single group of translators in the Aircraft Division.

C. THE TECHNICAL PUBLICATIONS DEPARTMENT TRANSLATION ENVIRONMENT

The basic issue of the contractual technical publications for one aircraft represents some 39 manuals and approximately 800,000 printed pages.

The major manuals are customized either to the airline fleet or to the aircraft and a revision service keeps them up to date.
With the multiplication of aircraft types produced, the sharp rise in aircraft sales and the rapid expansion of the in-service fleet, the quantity of technical publications shipped each year is constantly increasing. Over the last ten years the volume of technical publications shipped yearly has increased from 18 million pages in 1979 to 69 million pages in 1989.

The technical publications, whether AIRBUS or ATR, are produced in cooperation by the various partners on the basis of an industrial worksharing defined by the GIE. In both cases overall leadership for the technical publications has been awarded to AEROSPATIALE. As leader partner AEROSPATIALE is responsible for developing all the EDP (Electronic Data Processing) facilities required for the production of technical publications.

Considerable investments have been devoted to the research and development of high-performance software capable of coping with the ever-increasing volume of data to be processed and quantities of publications to be produced.

For the A320, the EDP systems used for the management, acquisition and finalization of the technical data have been totally redesigned to comply with the new requirements of ATA specification 100, which establishes rules for the presentation of the data, and to achieve greater flexibility in the production cycles.

The A320 is in fact the first aircraft in the world for which the Aircraft Maintenance Manual integrates the requirements of the ATA 100 AMTOSS concept (Aircraft Maintenance Task Oriented Support System) designed to improve the organization of the Maintenance Manual and to facilitate automated data retrieval. An open-ended system designated GIPSY (General Integrated Publications System) has been specifically developed to meet these requirements. With GIPSY the technical authors, assisted by numerous built-in aids, update the data files in real-time and it is possible to obtain customized outputs of the manual to the latest technical status as and when required. In developing these new systems, the Technical Publications Department has acquired high potential for innovation and participates actively in a wide range of projects aimed at improving existing Product Support services or creating new ones :

- Technical publications on optical disk (ADRES)
- Computer-Assisted Aircraft Trouble Shooting (CAATS)
- Order Processing Automated on-Line (OPAL)
- Technical publications stock and shipping management software system (APASHE)
- Onboard Electronic Library System (ELS)
- Maintenance Information Planning System (MIPS)
- On-line interrogation by the airlines of the manufacturer data banks
- etc....

This constant search for innovative methods of increasing productivity and enhancing the quality of service provided to the customers while reducing costs and production cycles has not neglected machine translation. At present, the Technical Publications Department disposes of a group of 11 full-time translators backed up by an equivalent number working as subcontractors. This group is responsible for all technical publications translation-related activities as well as the translation of various documents such as correspondence, technical reports, specifications, presentations, brochures, press articles, contracts, etc... issued or received by Product Support and a wide range of other departments within the Aircraft Division.

The possibility of using MT or CAT has therefore aroused wide-spread interest and numerous in-depth studies have been conducted to investigate the feasibility of integrating MT or CAT in this very specific environment. These studies have evidenced a certain number of problems that have to date rendered this integration impossible for technical and/or economic reasons.

2. PRACTICAL PROBLEMS ASSOCIATED WITH THE UTILIZATION OF MACHINE TRANSLATION

The utilization of MT/CAT inevitably raises a certain number of problems irrespective of whether the system adopted is a small system operating in a PC environment or a large system operating on a central computer or accessible on a subscription basis via an external network. These problems are mainly technical and economic but certain human and psychological problems, although of lesser importance in the decision process, should not be totally neglected. Two criteria generally determine the cost effectiveness of MT/CAT :

- the volume to be translated,
- the extent of human preparation/correction required to obtain a satisfactory result.

It is evident that the utilization of MT/CAT can only be envisaged if the volume of translation involved justifies the investments. Furthermore, any time spent by human translators in preparing source documents for MT/CAT (pre-editing) or correcting MT/CAT outputs to achieve the required result (post-editing) reduces the cost-saving capacity of the system.

In addition to these "universal" criteria, there are of course other more specific criteria that need to be taken into consideration such as the technical and economic aspects of integrating MT/CAT in a given EDP environment.

The decision as to whether or not to go MT/CAT is, therefore, based essentially on a comparative study between the constraints and problems inherent in the integration and utilization of the system and the estimated savings in terms of translation costs, leadtimes and personnel.

To date, the utilization of MT/CAT, in the very specific translation environment existing in the AEROSPATIALE AIRCRAFT DIVISION Product Support organization, generates a rather exceptional accumulation of constraints and problems that preclude a rational and cost-effective integration of the systems currently available on the market. Indeed only a very small percentage of the aircraft technical publications or the far from negligible volume of various other documents transiting through the translation office are suitable for MT/CAT.

A. TECHNICAL AND ECONOMIC PROBLEMS WITH MT/CAT

(1) Aircraft Technical Publications

(a) Volumes of data

The studies carried out to date have shown that, despite the impressive quantity of pages they represent, the aircraft technical publications contain relatively little text compatible with MT/CAT.

As has already been mentioned, the technical publications package for one aircraft comprises some 39 manuals.

Each of these manuals has been developed for a specific utilization in a specific context and complies with very strict rules imposed by industry standards such as the ATA specifications.

The content of certain manuals such as the Aircraft Wiring Manual and Aircraft Schematic Manual is essentially graphic.

In other manuals, the content is a combination of contracted text and complex layout.

In the Aircraft Trouble Shooting Manual the text is presented in the form of diagrams.

In the operational manuals, the text, generally written directly in the end-user language by specialized personnel, is characterized by highly-integrated graphics.

The Illustrated Parts Catalog, which is basically a series of illustrations with associated nomenclature, contains little text suitable for MT/CAT.

The conclusion of a detailed review of all the various manuals is that the cost-effective utilization of MT/CAT in the technical publications context is largely dependent on its utilization for the Aircraft Maintenance Manual (AMM).

Indeed the AMM is one of the rare manuals to contain text in sufficient volume and in a form that enables the cost-effective utilization of MT/CAT to be envisaged.

Th[illegible] divided into two parts :

- [illegible]etical part describing the various systems of the aircraft and their operation
- a practical part detailing the various procedures required for the maintenance of the aircraft.

It is during the preparation of the basic manual for a given type of aircraft, that, due to the considerable volume of text to be translated in a relatively short period, the utilization of MT/CAT is the most attractive. As an example, the A320 AMM for a given customer contains around 28,000 pages (≃ 17,000 text and 11,000 illustrations) and the total data bank around 40,000 pages (≃ 25,000 text and 15,000 illustrations).

When the basic content has been issued, the manual enters the revision phase during which it is updated at regular intervals (generally quarterly) to integrate changes relative to modifications embodied on the aircraft, variants specific to customized configurations or the correction of possible errors. During the first years in the life of a manual approximately 20 % of the pages are revised at each revision although the percentage of text actually new or modified is very much lower than this figure.

However, although these figures are quite impressive, it is important to note that, in today's context of European cooperation, the AMM is produced on an inter-partner worksharing basis and that this has a significant impact on the volume of translation.

In the case of the ATR, the AMM is officially issued in English and French. AEROSPATIALE is responsible for translating its contribution from French to English and the AERITALIA contribution from English to French.

As far as AIRBUS is concerned, the AMM is officially issued in English only. However, for the A320, AEROSPATIALE has separate contracts for the translation of the AMM from English to French.

For the official version in English, the other AIRBUS partners write their contributions directly in English. As a general rule, the AEROSPATIALE authors responsible for approximately 70 % of the manual, write in French during the initial production phase when time is short and the texts are long. However, once the manual enters the revision phase, the same authors tend to write modifications to existing texts or variants derived from existing texts directly in English. The French version of the manual is produced by retrieval of the texts written directly in French and translation of the partner contributions as well as the texts written directly in English by the AEROSPATIALE authors.

The quantity of translation involved in the production of the AMM is, therefore, less than the total volume of the manual would initially seem to indicate but remains sufficient to warrant an investigation into the utilization of MT/CAT.

(b) Integration of MT/CAT in the Technical Publications Production Process

The volume of translation involved in the production of the AMM having been judged sufficient to justify the utilization of MT/CAT, the next step is to study its integration in the production process.

One of the factors influencing the cost effectiveness of MT/CAT is whether or not the source text is available in a form that can be fed directly into the system.

At first sight, the ATR Maintenance Manual production environment seems particularly well adapted to MT/CAT in this respect.

Texts written in English are acquired in one file and those written in French are acquired in another. It would be relatively simple to integrate the MT/CAT between the two files to translate the source text (whether English or French) into the other language.

Unfortunately, however, for EDP reasons the ATR Maintenance Manual has the particularity of being acquired entirely in upper case. Tests performed on representative samples of the manual have shown that, due to the absence of lower case letters and more especially the accentuation, the results obtained with MT/CAT when translating from French to English are totally unacceptable.

In the case of AIRBUS and more particularly the A320, A330/A340 and future programs, the problems are different.

To comply with the latest requirements of the ATA 100 relative to the AMTOSS concept, to enable a more rational utilization of EDP data management systems and to prepare the way for the new Technical Publications media, AEROSPATIALE has totally redesigned the Technical Publications production systems.

For example, the ATA 100 AMTOSS concept requires that the manufacturer provide the airlines with a PMDB (Production Management Data Base). This bank contains data related to the planning and organization of maintenance extracted directly from the text of the Aircraft Maintenance Manual. The production of this data implies the integration in the text during acquisition of codes (tags) identifying the data for subsequent extraction. With the new Technical Publications media (such as optical disk) now being developed, these tags are also used to establish intra-manual and inter-manual links. It is important not only to establish links between data but to ensure consistency of data within a given manual, between manuals and with the placards on the aircraft and equipment. It was decided that, to avoid duplicating the acquisition of data with the risk of error this represents, wherever possible the data would be extracted directly from the data source file and automatically integrated in the text.

Another objective that largely influenced the design of the new production systems was that they should be capable of immediately outputting a customized manual fully updated with the latest known data without being subordinated to rigid revision cycles. To achieve this objective the systems were designed to enable the authors, whether in France, Germany, Great Britain or Spain to acquire their data on-line via terminals connected to the central data bank in the AEROSPATIALE TOULOUSE facilities.

In view of all these and other requirements, the new data acquisition system, designated GIPSY (General Integrated Publications System), was organized around file management software facilitating the transfer of data between files rather than conventional word processing.

In fact the system manages a certain number of different but interconnected files, each containing elements of the manual. The Maintenance Manual as such is the result of a finalization process which consists in extracting data from these various files, compiling the extracted data and presenting it in a form adapted to its future utilization by the end-user. This totally new Technical Publications production concept, based on the use of file management rather than word processing software, involves a very specific on-line data acquisition process. This process requires the use of 132 character screens and this, together with the presence in the text of numerous tags, considerably complicates the integration of MT/CAT.

English Version of the Aircraft Maintenance Manual

The possible integration of MT/CAT only concerns the AEROSPATIALE contribution (approximately 70 % of the AMM) and more precisely that part of the AEROSPATIALE contribution written in French. It is evident that the integration of MT/CAT in GIPSY must not have an adverse effect on the performance of the system as a whole to the detriment of the other partners and those French authors writing directly in English. It should be noted that, as the official language is English, GIPSY imposes English as the source language and numerous aids (such as the automatic generation of standard sentences and automatic call-up of technical designations) have been built into the system to prompt the authors to write directly in the system in English. The acquisition of English as the GIPSY source language requires that the French-to-English translation phase be located upstream and independent of GIPSY. This implies a preliminary acquisition of a pre-edited French text for the MT/CAT process and a second acquisition under GIPSY of the translated text after post-editing to correct the translation and to restore the coding specific to GIPSY, etc.

This process considerably retards the availability of data and is contrary to the on-line update principle. Furthermore, it is difficult to clearly define responsibilities for the various phases of the process.

Initial write-up and integration of tags to identify significant data are the responsibility of the authors whereas the pre-editing and especially the post-editing fall rather under the responsibility of the translator.

The utilization of MT/CAT in conjunction with GIPSY for the production of the English version of the AMM therefore imposes serious drawbacks which are contrary to the basic philosophy of GIPSY.

The complete process which involves acquiring the data twice, is lengthy, complicated and not at all cost-effective.

This is particularly the case when the manual is in the revision phase during which most of the work involved in updating the manual is such that the French authors prefer to write directly in English. Finally, one other factor complicates the use of MT/CAT for the production of the English version of the manual. The ATA 100 now requires that the AMM be written in Simplified English. Simplified English is a controlled language specifically developed by AECMA (Association Européenne des Constructeurs de Matériel Aérospatial) and the AIA (Aerospace Industries Association of America Inc.) for aircraft maintenance documentation. Simplified English consists of a limited vocabulary and a set of writing rules for using that vocabulary. The limited vocabulary (approximately 800 words) includes, verbs, prepositions, conjunctions, adjectives, adverbs and nouns. In this limited vocabulary, a family of synonyms is represented by only one of its members. For example "start" is used instead of "begin, commence, initiate or originate". Also, as a general rule, the words in this limited vocabulary have only one meaning. For example "fall" is used to indicate the idea of gravity and not the idea of decrease in quantity. Finally the words in the limited vocabulary can only be used as the part of speech indicated in the dictionary. For example "check" can be used as a noun but not as a verb.

As for the writing rules, they have been developed to make the written message easier to understand by users of the manual whose first language is not English. Sentence length is limited to 20 words, verbs are used in three tenses only (present, past and simple future), noun clusters are broken down, the passive voice is to be avoided, articles must be used, etc....

Although Simplified English is easier to understand for the users of the manual for the Manufacturer it constitutes a new constraint. To comply with the requirements of Simplified English it is often necessary not only to replace unapproved words by approved ones but to completely reformulate the initial idea. For example, "switch" being an approved verb, "switch on NAV1" has to be reformulated to something like "set the NAV1 switch to the ON position". This reformulation is fairly simple for the authors writing directly in English. However, for the translator, the reformulation may require additional information not contained in the initial sentence. For example, "Action on the ENG/FIRE pushbutton switch arms the squibs" cannot be transformed into good Simplified English without specifying what "action" is required (push or release ?). Whereas the human translator can cope with such problems, if necessary by contacting the author, this is not the case of MT/CAT.

The requirement to write in Simplified English is therefore another argument against the use of MT/CAT for the production of the English version of the AMM.

French Version of the Aircraft Maintenance Manual

The French version of the manual is produced by extraction and translation of the applicable texts existing in English in the GIPSY files.

The translation of the AEROSPATIALE contribution is facilitated by retrieval, where available, of the drafts written in French.

The ideal situation would be to transfer the contents of the English files to the French files via an integrated MT/CAT stage.

Here again, however, the specific formats of the GIPSY files raise problems of compatibility with existing MT/CAT systems. The development of programs capable of automatically converting the contents of the GIPSY files to and from an MT/CAT compatible format was envisaged but had to be abandoned for cost reasons (cost of program development and subsequent processing time).

Due to these interfacing problems and as for the English version of the manual, the only possibility is to keep the MT/CAT system independent of GIPSY. This implies obtaining outputs of the English texts, pre-editing and acquiring them for MT/CAT, post-editing the results and acquiring them in GIPSY format.

This is obviously a lengthy process which cancels out the advantages of using MT/CAT.

Furthermore, GIPSY offers the translator working directly in the system aids similar to those available to the authors (automatic generation of standard sentences, technical designations, etc...).

To date, therefore, the constraints imposed by the large technical publications production systems on the one hand and the MT/CAT systems on the other are such that it is difficult to envisage a rational and economic integration of the two with current technologies.

The use of MT/CAT has also been investigated for the production of another technical document called Service Bulletin (SB). An SB is a self-contained document describing a modification and containing instructions for embodying the modification on the aircraft or an item of equipment. An SB generally comprises between 5 and 200 pages and annual production is currently around 13,000 pages.

Here again, the specific nature of these documents, the possibility of retrieving standard sentences, the obligation to strictly comply with technical designations and the requirement to write in Simplified English cancel out by the pre- and post-editing involved any advantages that could result from the utilization of MT/CAT.

Indeed this is confirmed by the fact that one of the AEROSPATIALE translation subcontractors, who is also the agent in France for a well-known CAT system, finds it more economical to translate SBs "manually". Furthermore, the Airbus SBs will soon be produced using GIPSY.

(2) Technical Problems Associated with the Translation of Miscellaneous Documents

The Technical Publications Dept. Translation office is not only responsible for the translation of the manuals but also of a large quantity of miscellaneous documents for the Product Support and other Aircraft Division Directorates. These documents can be divided into two categories :

- outgoing documents, generally written in French and translated into English,
- incoming documents, generally received in English and translated into French for internal utilization.

These documents represent the translation of some 10,000 pages annually which is sufficient to envisage the utilization of MT/CAT.

Outgoing Documents

The source texts generally arrive in the translation office in the form of hand-written rough drafts. When the author is distant from the translation office these drafts are often sent by fax.

It is also worth mentioning that these drafts are written by authors who know that their texts are going to be translated and if necessary re-organized into a logical presentation, completed or corrected.

The utilization of MT/CAT requires firstly that the source texts be acquired into the system and secondly that the source texts be of a quality compatible with a satisfactory result. MT/CAT does not escape the rule applicable to all computer systems : "garbage in, garbage out". Investigation into the outgoing documents has shown that, in the majority of cases, the time required to prepare the texts for MT/CAT and to post-edit the results often exceeds that necessary for a human translation.

Furthermore, an ever-increasing number of these documents and especially correspondence addressed to the partners (memos, meeting reports, etc...) is being written directly in English with the assistance of or a quick check by the translators when deemed necessary.

The translation of outgoing documents is tending to be limited to those texts requiring special attention (contracts, specifications, technical reports, press articles, brochures, etc ...) and which would require careful post-editing if MT/CAT were used.

Incoming documents

The majority of incoming documents arriving in English are used directly as such without being translated into French. The role of the translator is often limited to providing verbal confirmation of the correct comprehension of certain specific points. There is, however, a relatively small number of documents for which a written translation is required.

To enable the utilization of MT/CAT for the translation of these documents they must first be transferred to an EDP media. The simplest and quickest way of doing this is to use a scanner but this is only possible if the document is of sufficient quality : typed, no stamps, no annotations, no folds or marks from photocopy machine, etc ...

Unfortunately such quality is rare and, therefore, most of these documents would have to be re-typed into the MT/CAT system. In most cases a certain amount of pre-editing would be required and glossaries would have to be updated. The post-editing, however, could be "à la carte" ; in some cases the draft translation with little or no post-editing would be sufficient, in other cases fine post-editing would be required to achieve the desired standard.

As the majority of incoming documents are fairly well adapted to MT/CAT, the justification of an MT/CAT system would depend essentially on the volume of translation concerned and a reduction in translation lead times. The current price of subcontracted translation outside PARIS and notably in the TOULOUSE area is close to if not lower than the price of MT/CAT translation. The cost-effectiveness of MT/CAT in this context is therefore largely dependent on there being a requirement for a large quantity of translation with a minimum of post-editing.

These conditions cannot be met without increasing the number of potential users by becoming a central server for the whole of the Aircraft Division or even the Company.

B. HUMAN AND PSYCHOLOGICAL PROBLEMS

The arrival in a department of new tools and methods always gives rise to a certain apprehension especially when computerization is involved. This apprehension is rapidly dissipated if the new tools and methods prove efficient and improve working conditions.

In the case of MT/CAT, the apprehension of the translators is often amplified by the more or less justified impression that the system constitutes a direct rival rather than just another tool at their disposal. Indeed, at the very idea of MT/CAT, the majority of translators see themselves being subordinated to the machine, deprived of the creative aspects of their work and reduced to trivial tasks peripheral to translation proper, such as pre- and post-editing, updating glossaries, etc ...

These fears, although often exagerated, are not always totally unfounded.

It cannot be denied that, in the minds of many, translators represent a source of extra costs and MT/CAT a means of limiting them. It is obvious, therefore, that the acquisition of an MT/CAT system must result in significant savings in terms of translation costs and lead times, either by a reduction in the number of translators or by an increase in productivity.

The utilization of MT/CAT can, therefore, place the translator in an ambiguous situation where the notions of reduced costs and lead times conflict with that of "quality".

This situation is of course mainly related to the post-editing of the draft translations output by the machine. Translation, even when technical, can be highly subjective and this subjectivity can give rise to a certain frustration during the correction of translations produced by someone else ; especially if this someone else is a computer. Faced with a draft translation from the machine, the translator no longer feels totally responsible for the quality of the translation but obliged to accept a compromise.

Some, over-conscientious, translators will tend to completely rework the translation to the detriment of cost-effectiveness. Others, not sufficiently conscientious or influenced by the machine outputs will tend to accept the draft translations as such to the detriment of quality.

There is a general risk that, through too much compromise, the translators end up losing interest in their work.

These general problems may be aggravated by others of a more specific nature. At AEROSPATIALE, most of the translators have received a litterary education but work in environment which is hyper-technical. They must, therefore, make considerable efforts to acquire not only the appropriate technical vocabulary and style but a wide technical knowledge of their field of activity.

The technical authors, on the other hand, have generally received a purely technical education and must make considerable efforts to acquire the art of writing correctly. The result of this is that a natural balance develops between the author and the translator, each, as it were, compensating for the deficiencies of the other.

The arrival of an MT/CAT system could upset this balance. The translators may have problems adapting to the new situation or may even be obliged to change activity. Indirectly, however, there may also be problems for the authors whose texts are not always suitable for direct translation by a machine. In order to reduce to a minimum the problems related to the integration of an MT/CAT system, it is essential to prepare a detailed specification and to associate the translators and the authors in the preparation of this specification. Without this specification, there is a serious risk of acquiring a system that is not adapted to the requirements. Used in a rational manner, in a well defined and controlled context, MT/CAT could liberate the translators from tedious routine translations and leave them more time to concentrate on the translations that are ill-adapted to MT/CAT or that require special attention.

The fact is that many of the problems associated with MT/CAT are due not to the system itself but to a poor evaluation of the needs and a false idea of the capabilities of the system.

Only too often there is a certain reluctance to accept the studies carried out by the translators themselves on the grounds that they are too anxious to protect their profession and refuse progress. In reality, however, translators in general and the AEROSPATIALE translators in particular have, by the very nature of their work, developed a considerable capacity for adaptation. They are fully aware of the fact that it is better to actively participate in and thereby influence progress rather than to obstruct it artificially and finally have it imposed upon them.

It is therefore essential that the translators play a predominant role in the preparation of the specification relative to MT/CAT and that they participate in the decision as to the eventual choice of a system. It is thus possible to avoid problems arising from a poor understanding of the systems, of what they can do and what they cannot do.

Either the MT/CAT system is adapted to the real needs of the given context, in which case there are few problems whether psychological or otherwise, or it is not, with all the consequences that this implies.

3. CONCLUSION

There is currently a wide range of MT/CAT systems, both large and small, on the market capable of providing valuable services in well defined contexts.

Too often, however, the acquisition of an MT/CAT results in failure. In most cases this failure is not due to the system itself but rather to the fact that it is not adapted to the specific user requirements either because these and the integration of the system were not sufficiently analyzed or because the capabilities of the system were over-estimated for the given application. Several "general purpose" systems, after a commercially successful period have seen their sales drop dramatically. Other systems, designed in a given context to meet clearly defined requirements continue to give full satisfaction and in these cases the research and development is being actively pursued to further increase their performance. At the present time, it would seem as though MT/CAT is going through a period of transition where it is important to learn from the errors of the past to better prepare the future.

At AEROSPATIALE, the aircraft technical publications are also entering a period of transition and substantial investments are being devoted to preparing the future. The technical publications systems are being totally redesigned to enter the era of fully computerized transmission of technical data. The major preoccupation is to develop open-ended systems compatible with the new concepts of data management, organization, transmission and utilization. AEROSPATIALE is currently capable of providing technical publications on CD-ROM and is actively working towards the direct consultation by the Airlines of the Manufacturer data bases. An advanced studies group is starting to investigate the yet unexplored field of "intelligent" graphics. All these developments have and will continue to have repercussions on translation and the needs for and utilization of MT/CAT.

In the constant search for new ways of reducing production costs and lead times while increasing productivity, MT/CAT was immediately seen as a valuable means of contributing to these objectives. However, this position has now been modified mainly because of the problems involved in integrating MT/CAT in the new technical publications production systems.

The current developments do not facilitate this integration and, in fact, the whole technical publications environment is moving so fast that it is difficult to precisely define needs. It is, however, certain that an eventual MT/CAT system would have to be capable not only of adapting to this environment but also of evolving with it. The MT/CAT systems available today are best suited to fairly long texts with conventional layout.

However, the current trend with aircraft technical publications is to break down the texts, whether descriptive or procedural, into small highly-coded documentary units of just a few lives. These documentary units contain little "free" text but include numerous codes for calling up precise items of technical data or standard terminology/sentences from associated source files. It must be stated that these new production systems were designed to satisfy a certain number of essential requirements and that the integration of MT/CAT was not one of them. Due to the complexity of this integration and the risk of having an adverse effect on overall system performance notably to the detriment of the Partners who do not require MT/CAT capabilities, the MT/CAT solution has, for the moment at least, been abandoned. Efforts are now being concentrated on getting the technical authors to write directly in English. It must be admitted that the majority of texts in the Aircraft Maintenance Manual do not present any major linguistic difficulties and that the author aids, developed with the assistance of the translators and built into the system, greatly facilitate this. The translators are of course available for linguistic assistance or the translation of more complicated texts when required. This method of working has the added advantage of being consistent with the basic philosophy of the new technical publications production systems which calls for immediate availability of data through on-line acquisition by the authors.

As far as technical publications are concerned, therefore, the role of the translator is evolving towards more and more terminology as opposed to translation as such. Indeed the translators are involved from the earliest stages of a program in preparing the terminology specific to that program (equipment and system designations). This terminology will be used on the design drawings, on the aircraft itself and throughout the technical publications. The role of the translator, therefore, is increasingly to initialize author aids integrated in the technical publications production systems.

There remains, however, a wide range of other documents for which the services of a translator are still required although, here too, there is an ever-increasing tendency to write or use the more simple documents directly in English. A significant proportion of the miscellaneous documents requiring translation is compatible with MT/CAT but the volume they represent within Product Support alone is not sufficient to warrant the acquisition of a sytem. This could only be justified by offering the services of the MT/CAT system to a wider population within the Company.

Despite the difficulties currently encountered with MT/CAT, the translators in the AEROSPATIALE AIRCRAFT DIVISION PRODUCT SUPPORT organization continue to follow and participate in the development of various MT/CAT systems. They have prepared a specification for an MT/CAT system adapted to their specific working environment but, to date, no system has been found,that meets the requirements.

A study group organized by these same translators is tending towards the definition of what has been termed a "translator workstation" rather than towards conventional MT/CAT. These workstations, better adapted to the diversity of tasks performed by the translators, would be connected in a ring to a server and dispose of a certain number of shared translation aids such as word processing, glossary management software, reference documents on CD-ROM and could, if applicable, integrate an MT/CAT system. These workstations would be fully integrated in the Company EDP and office automation environment :

- direct access to the EDP network for consultation and update of the files managed by the mainframe computer and particularly those related to the technical publications.
- direct access to the office automation network to facilitate the transmission of documents between the translators and their "customers" and to access reference data bases connected to this network.

The translator workstation will probably constitute an intermediate solution for the short and medium term but it is difficult to forecast what the long term situation will be. At present, the fact that English is considered as the international aeronautical language and used as the official language in the AIRBUS and ATR GIEs combined with the fact that the vast majority of the AIRCRAFT DIVISION products are sold to export has prompted AEROSPATIALE to make considerable efforts to produce and use documents directly in English. However, the recent changes in the European political scene and their repercussions on the international commercial landscape will inevitably have an impact on the demands for translation.

There is no doubt that new generation MT/CAT systems, with increased performance and new capabilities will have an important role to play in tomorrow's world of international communications. It is, therefore, essential, despite the difficulties encountered today, that Companies whose activity is largely dependent on international commerce and therefore international communications continue to participate in the development of MT/CAT systems. This is the only way to ensure that the new generation systems will be really adapted to their requirements.

END USERS' NEEDS

by

Thomas Schneider
Siemens AG
DI AP 341
D-8000 Munich 83

To assure international communication the introduction of machine translation systems is unavoidable. To be of use in practical applications, however, a system must fulfill the criteria of operability defined by the end user. In this context, two different applications must be contrasted, namely automatic translation with the aim of gathering information for internal applications on one hand and automatic translation with the aim of producing publications for external recipients on the other hand. In the first case, throughput and coverage of the system lexicon are most important while preservation of layout and format information are secondary. In the latter case, a much higher translation quality is required to ensure user acceptance. To be effective in an office environment, additional aspects become relevant: user interface, integration with other office systems, efficient lexicon update and postediting tools. Necessary for all types of applications is an intensive end user training and continuing support from specialist consultants.

In the light of the explosion of knowledge and the necessity to gather and exchange information across national and linguistic boundaries, the introduction of machine translation has become inevitable. Human translators may feel threatened in their job security but such fear is usually caused by a misunderstanding of what a machine translation system can and cannot do. A machine translation system, even the most powerful one available today, will not replace a highly qualified human translator. There is no linguistic theory in sight which would permit the complete and unambiguous analysis or generation of a single natural language. In other words, a human revision of machine-translated texts will always be necessary, and certain types of text will by definition be reserved for human translation, namely any text in which nuances of style need to be preserved or in which the meaning is hidden "between the lines". These types of text unsuitable for machine translation include not only literary works but also political speeches and quarterly reports.

For the translation of texts conveying factual information such as technical documentation, scientific abstracts or fact sheets a machine translation system is a powerful tool able to increase a translator's productivity by several factors - provided the system is designed with the translator's requirements in mind. To define the needs of an end user we need to differentiate two possible applications, the use of machine translation to gather information for internal purposes on one hand, and the use of MT with the aim of producing publications for external recipients.

Machine translation for information gathering If machine translation is used for purposes of information gathering it is likely that the texts to be translated come from various heterogeneous sources and cover a wide range of different topics. For an MT system to operate effectively it must contain a very large lexicon covering many different subject fields. However, one must keep in mind that it is by definition impossible to incorporate a "complete" lexicon for all applications. The general vocabulary of a language like English may amount to perhaps 300 000 entries. The sum of the concepts in the various sublanguages, e.g. in medicine, chemistry, data processing etc, by contrast is estimated to be in the area of 30 to 50 million. The vocabulary within specific subject fields increases much more rapidly than that of the general language, and it would be a hopeless task to try and keep up with the lexical change in all subject fields. So even for purposes of information gathering, a kind of specialization is necessary.

When comparing the lexicon size of different machine translation systems one should not take a vendor's figures at face value. The structure of a lexical entry may differ greatly from one system to another. One possibility is to list all complete word forms as seperate entries. For a language without major inflection like English this might not be too disadvantageous. For languages like German or French however such an approach would inflate the lexicon by a factor of ten or more, without increasing text coverage at all. For the German verb "bestehen" for example there would have to be seventeen separate entries.

Another approach will carry separate entries for each word stem. The English verb "go" would be listed under three entries, for "go", "went" and "gone". Again, this will inflate the number of lexical entries. More modern systems will have just one entry per headword and generate all inflected forms by reference to grammar rules or by consulting morphological tables. Another aspect to be considered is the treatment of compound words. Some languages such as German can form new terms by linking existing words. In the majority of cases, these compounds can be translated into English on the basis of the individual components. A German term like "Plattenspeichersubsystem" will translate nicely as "disk storage subsystem". Provided the machine translation system has grammar rules which are able to analyze such compounds, a lot of lexical entries are superfluous.

When dealing with large numbers of documents from heterogeneous sources it would be advantageous to automatically identify the subject field of the text. At present, no machine translation system seems to have this feature. The same word of course will denote entirely different concepts depending on subject field and require a different translation. A "trunk" may be a part of a car or a communication line, not to mention a suitcase-like container or an elephant's proboscis. To ensure an adequate translation, the MT system needs to be geared to the relevant subject field. If it cannot be done automatically the human translator needs to set the "bias" manually.

If the documents to be translated come from many different sources it is highly unlikely that they all adhere to the same formatting conventions. It is more probable that a variety of usually incompatible editors and word processors are used, and that a fair percentage of the texts may even be composed on paper. That greatly diminishes the chances of being able to process the material in machine-readable form. For the application of machine translation of sundry documents for purposes of information gathering, the installation of a font reader should prove economical. Having a staff of typists input the texts received is not just expensive but introduces errors. Font readers, to be sure, are not perfect but some of the newer models are able to handle a variety of fonts and have "learning" capability, i.e. the ability to adapt the recognition to specific features of the document.

One of the great problems in the production of multilingual documents is the need to preserve the format of the original. That aspect is fortunately of less concern in information gathering. It is usually sufficient to extract the relevant text portions from the document. Only in rare cases, such as the interpretation of flow charts and tables, will the page layout be of sufficient interest to warrant its reformating. As a rule, a translation of the running text will be enough of an aid for a specialist to understand the content of the document. Should the automatic treatment of the document prove inadequate a human revision or even a completely new human translation might be added.

Collecting information from foreign-language sources involves several parameters. Usually a very large amount of text needs to be translated, and the rapid accessibility of the information is of great importance. As the information is usually utilized by specialists who know the subject fields well the stilistic quality of the translation is of lesser consideration. In practical terms this means for an end user that the machine translation system has to be able to process a lot of text very rapidly. If a thousand documents per day need to be translated a sophisticated MT system with superior translation quality but insignificant throughput would be useless. However, as was pointed out earlier, it would be impossible to have a single translation system dealing adequately with all types of text and all subject fields simultaneously. Therefore it might be worth a consideration to run separate systems for different document types, and with this kind of specialization slower systems may prove to be adequate in throughput. Rough translations for the purpose of information gathering do not require the degree of linguistic sophistication in the system that is needed for texts to be published. Very often a local grammatical analysis, i.e. one based on the analysis of phrases, may suffice. The sentence translated at the phrase level might make a human translator throw up his arms in despair but the rough content might still be understandable to an expert in the field of the text. A note of caution should be added: a phrase level translation could in this context be adequate for an English text, or for any other language that has a rigid word order. It would probably not work for languages in which phrase elements are not necessarily contiguous, e.g. German.

The quality of a translation cannot be measured in percentage points. To define a quality level which is acceptable in the area of information gathering is impossible. This would hinge on aspects such as the type of source language, the familiarity with the subject matter on the part of the reader and the degree of precision required. These factors may vary greatly from one application to another.

One should expect a machine translation system intended for the "quick and dirty" translation of large volumes to run on main frames, perhaps with access from various sites. However, aside from the fact that general purpose mainframes are notoriously ill-suited for the processing of natural languages, there is another factor to be considered. Wide accessibility of the information may be intended in some cases. In other environments, the information may be classified and may need to be protected against illegal access. Such protection may be difficult if the translation system resides on a general purpose mainframe which is widely accessible. The problem is compounded if the mainframe is integrated in a network. A possible alternative would be a stand-alone machine translation system which could more easily be protected and which could be made radiation-free.

One of the goals in connection with machine translation for such purposes is the automatic processing of the facts contained in the documents for storage in a data base. There are currently several projects under way with this goal, e.g. at Siemens in Munich, but besides some severe theoretical problems there is the enormous expense of quantitative work load. No quick solution is to be expected here.

Machine translation for the production of publications If the gathering of information requires primarily speed of throughput in the translation process, the use of machine translation for the production of documents to be published for an outside world demands a complete package of solutions to be viable.

First of all, it is usually a requirement that the target text be of high quality; in some cases it is required that the translated text should not be recognizable as a translation. This presupposes human revision of the machine translated text. No system, no matter how sophisticated, could fulfill this requirement. This, however, is not to be misunderstood as a naive notion that all systems are created equal. For a machine translation system to be used effectively, the human translators have to accept it as a tool, and one of the prime requirements is high translation quality. If translators have to correct too large a percentage of the translations proposed by the machine they will view the system as a burden rather than an aid to productivity. And if the system is not acceptable to the translators no gain will be realized. On the contrary, personnel problems might develop.

The achievement of high quality translations presupposes a thorough linguistic analysis. No word level or phrase level analysis can provide the basis for a plausible interpretation of a sentence; the minimum requirement is an analysis that takes all elements of a sentence into consideration. Such an approach may be "expensive" in terms of computing power but in the area of natural language processing there is no choice. The many ambiguities in the words of a sentence cannot be resolved by minimalistic local analyses. Natural languages are - contrary to the assumptions of the past - not finite systems. Therefore it is important that an MT system has a grammar which does not just list a finite set of legal grammatical structures. Otherwise the next document, written in a slightly different style, may turn out as unintelligible gibberish.

Translation quality depends to a large extent on the power of the grammar, but equally important is the information contained in the lexicon. And here it is an indispensable requirement that the translators working with the system have complete access to the lexicon, to be able to enlarge and adapt it.

No matter how large the lexicon supplied by the vendor, there will always be the necessity to update it. Specific applications in defined subject areas demand the use of specific terminology usually not contained in general lexicons, and in industry very often company-specific terminology takes precedence over more general terms. The lexicon of a powerful machine translation system contains a lot of grammatical information which is used for the analysis or generation of a language in unison with the grammar rules. Updating the lexicon does not just mean adding word pairs in two languages but adding morphological, syntactic and semantic information. A system might need to know that "rely" is a verb, that it occurs with the preposition "on", that its inflection is regular etc. Keying in all this information may be quite cumbersome if it is not supported by tools. In one commercially available system, coding a new term takes a full-fledged linguist half an hour. Such figures make a system excessively expensive. In another system (METAL) the coding of lexical entries is supported by an integrated expert system so that new subject fields can be added to the system lexicon with minimal expense.

But not only the coding of lexical entries needs to be open to the end user. An acceptable translation quality can only be achieved if the translation is geared to a specific subject field. Most systems nowadays provide a framework for subject-specific lexicon modules so that in the translation of a given text highest priority is given to the transfers contained in the most specific module. It is important, however, that an end user can not only fill existing slots but that he can define the structure of lexicon modules himself. There is no such a thing as the universal classification system, and a certain user in the field of chemistry may have entirely different requirements for his lexicon structure than a user in the field of civil engineering.

If machine translation is used to generate publishable documents the integration into an office environment is of utmost importance. The best translation quality is wasted if there is no smooth-running sequence of steps from the original to the target text. First of all, the original text needs to be imported from an external source into the machine translation system. There have to be physical means for this task, floppy disk drives, tape units and possibly ethernet connections to the word processing systems on which the originals are composed. It is important that besides the text, all graphics and other non-linguistic material are preserved. This is not a trivial task as we are still faced with a multitude of imcompatible editors and word processors, all of which seem to encode graphic information differently.

It would be uneconomical to manually extract the text portions to be translated. So there has to be a set of programs to automatically indentify the translatable text portions and separate them from the non-translatable material - which may constitute more than 50 % of a page. This information needs to be preserved so that it can be used to reconstitute the format and layout of the original page after the text has been translated and revised. As it does not seem likely that such programs will work perfectly on all types of documents, the end user must have the capability to override the automatic process and edit the various intermediate versions.

In the early days of data processing the user had been expected to adjust to the formalisms demanded by a system. Fortunately, some progress in this area has made the end user's life easier. Most machine translation systems nowadays offer a menu-driven interface which is easy to operate even by translators who are not knowledgeable in the field of data processing. By clicking on pre-structured command lines, translators can run the system without endangering its integrity. As the revision of a machinetranslated text requires different editing functions from those needed in general word processing most systems offer specialized editors. Typical functions include global replace, transposition of words or phrases etc. If a system is not an insular implementation of a machine translation system but resides as an add-on module to standard hardware and software it can of course be used for other purposes as well, an aspect which may figure in a cost/benefit analysis.

Machine translation systems are not self-explanatory. Even if end users are supplied with adequate documentation they need intensive training by experts. It is not sufficient to be shown the surface handling by a salesman. In order to understand the operation of the system properly, users need to be shown the system structure, the interrelation between grammar and lexicon. Only then will they be able to avoid costly errors in updating their lexicon. A machine translation system should not be a black box but a transparent tool for the translator. As updating the lexicon is one of the major tasks for translators, a fair amount of the training period should be spent on all aspects of lexicon work. The training should not only stress the linguistic parts of lexical entries but should include an analysis of the end user's subject areas and terminology structure as well. Finally it is important that an end user is shown the methods of postediting a machine- translated text as this differs from revising a "human" translation. Prototypical systems are usually not fit for productive applications. End users need long-term support in the operation of their system, both in hardware and software maintenance and in organizational and linguistic consulting. As more progress is made in the field of linguistics it would be wise to choose a system with a modular structure that can incorporate future developments be they additional languages or more powerful linguistic components.

ASPECTS OF MACHINE TRANSLATION IN THE UNITED STATES AIR FORCE

Dale A. Bostad
Foreign Technology Division
United States Air Force

Machine translation is used in the USAF to translate technical literature in a wide variety of disciplines to support studies that assess the capabilities and research of the Soviet Union and other countries. Two types of machine translation are used: partially-edited machine translation with hard-copy printing of the text, and raw machine translation for rapid information scanning of material at a terminal.

A special software program is used for rapid post-editing of texts. Potential trouble spots and ambiguities are intercepted and corrections are made by post-editors.

Raw machine translation for gisting large volumes of information has proven to be an effective tool for analysts and researchers. Statistics indicate an ever-increasing use of machine translation for rapid information scanning.

It is the purpose of this paper to describe the use of machine translation in the USAF, to describe the product, and to indicate future developments in the area of rapid information acquisition.

Machine translation is used by the USAF to support studies carried out by scientific and technical researchers who need to stay abreast of foreign developments in a wide variety of technical fields. Most of the translations produced come from open-source literature. A researcher has a broad base of information at his disposal, including extracts, abstracts, cover-to-cover translations of foreign journals produced by publishing houses, and other studies in the field. Machine translations are simply another source of information.

There are two general ways that machine translation is used in the USAF. The first use is partially-edited machine translation where a trained linguist massages or corrects the raw machine translation, bringing it to a higher degree of accuracy and readability. The second one is the use of raw machine translation for information scanning. These two applications will be discussed in some detail.

Partially-Edited Machine Translation

The standard product of the Directorate of Translations of the Foreign Technology Division (FTD) is partially-edited machine translation. Between 50,000 and 60,000 pages of Russian text are translated each year by the US Air Force Systran Russian system. This is a batch MVS system, often translating 10,000 sentences in 40 minutes (clock time). The system operates in both a classified and an unclassified configuration; however, the majority of translations are from open-source literature and they are translated on the unclassified system. These translations produced meet the standards of adequacy for the users of the product. Machine translation has gained wide acceptance by users because it provides rapid turnaround of information and the translations are technically accurate in a wide range of technical disciplines and are readily comprehensible to a subject-area specialist.

A partially-edited translation is produced by scrolling through the entire translated text on a video-display terminal. However, only segments of the text are actually post-edited. In fact, only about 20% of a given text is carefully looked at. What is to be edited is determined by a software program called EDITSYS. The functioning of this program will be examined in some detail.

EDITSYS

EDITSYS is a program called at the end of the translation procedure that serves to direct the post-editing, i.e., tells the editor exactly what to look at and edit. The program identifies trouble areas in the system that need review and intercepts these conditions. As stated, large chunks of text go through unscrutinized without editing. This means that we rely heavily on the efficacy of the linguistic algorithms and our large dictionaries. To write a program like EDITSYS one must have considerable knowledge about the strengths and weaknesses of the system and the programming expertise to highlight the weaknesses for review.

The program itself is a module that allows us to go in and test at the bit/byte level the final analysis area of sentences. Virtually all of the linguistic macros in the system can be used for testing. When a given test condition is met the program generates a full-width line of a certain character in front of the condition, and this line is interspersed in the text and displayed on the screen. As an editor scrolls through the translated text he halts whenever a flag line appears, and makes an editing decision. If no editing is required he continues on to the next flag; otherwise he corrects the error. Post-editing is limited to the immediate environment around the flag. A skilled editor can edit 15-20 Russian pages an hour using this technique.

Flags are generated by EDITSYS to check the following situations:

1. Not-found words. All legitimate not-found words or words incorrectly input are flagged. True not-found words are now relatively rare, since the dictionaries contains 200,000 entries.

2. Acronyms. All acronyms are checked to see if their expansions are correct. Thousands of acronyms are expanded in the dictionaries, but those of three characters or less require close scrutiny.

3. Rearrangement. Byte 144 indicating rearrangement is flagged. Approximately 20% of Russian sentences are rearranged with an accuracy rate of 90%. One sentence out of ten must be edited where words or phrases are moved into incorrect slots.

4. Contiguous slashed entries. There are several thousand slashed entries in the Russian system, and when slashed words in English occur next to each other smooth reading of the text is impeded. The most frequent occurrences are adjective + adjective, adjective + noun, and noun + noun.

5. Spurious "good" terms. These are words that have been typed incorrectly but which match up against the dictionary. Examples are *BOLE* instead of *BOLEYE*, *SOYA* instead of *SLOYA*, and *BIT'* instead of *BYT'*.

6. Uncertainty code. Byte 57,04 is tested. This uncertainty code is turned on in certain homograph routines at the point where the logic becomes tenuous, there is no statistical evidence for one dictionary default over another, and in fact resolution is a toss-up.

7. Problem words. There is a flag generated for certain problem words (about 40 in number) which the system has not been able to resolve with sufficient accuracy. This category is fluid; as routines or expressions are developed for these words they are no longer flagged. Of course, new conditions or words also arise which require flagging.

Raw Machine Translation

Three years ago the USAF developed a new application of its machine translation system which we call interactive machine translation. This system gives all users individual access to machine translation at their own terminals. It is now available to users on approximately 1400 PC's within the Foreign Technology Division. This is raw machine translation without the mediation of translators.

The system is designed so that a user can rapidly determine the significance of the material he wants translated and weed out extraneous information. It is best used for rapid translation of titles of books, tables of contents, captions under tables and graphs, and individual sentences and paragraphs. However it can also effectively be used to translate complete short articles and to get back a rapid translation instead of going through the sometimes time-consuming operation of routing translations through the formal bureaucracy. One very effective use of the system is for gisting a large book, that is determining the significant parts of a book and then routing this material through the normal translation procedures. For example if a user has a 350 page book, the system might be used to determine that only Chapters 3, 7, and 12-15 are really pertinent to his research. Obviously, by using such a tool there can be tremendous cost savings by not translating irrelevant material.

Computer Environment of Interactive MT

The first thing that had to be done to develop interactive MT was to reconfigure the Systran systems to run under IBM VM/CMS (Conversational Monitor System) operating system. The seven-step traditional procedure was reduced to a single step and the two IBM sorts were eliminated. In their place, random access searching was used in main dictionary look-up. Random access lookup of words is very efficient when processing shorter files. These were the changes required as far as Systran was concerned.

The system was then loaded on an IBM mainframe on a Systran disk. When a user is connected to the mainframe, all he need do is type in the command SYSTRAN and he is automatically linked to the Systran disk on the mainframe. The user has his own virtual machine running on the host. This means that he commands nearly the equivalent computing power as if he had the full resources of the mainframe at his disposal. Thus, on either the unclassified or classified system, all the user has to do is type in the command SYSTRAN at the CMS prompt and he is ready to execute a translation session.

Interactive MT Menu

The interactive menu was written by two FTD systems programmers using VM/CMS and XEDIT and REXX macro languages. The primary consideration was to make the menu simple to use and as short as possible, i.e., user-friendly. I will briefly describe the menu and options available. After typing in SYSTRAN the first panel appears, displaying

SYSTRAN

FTD's Interactive Language Translation System

on the screen. Specially defined function keys at the bottom of the screen then direct the user to proceed to the next menu, or exit. All subsequent menus have dedicated function keys, explained at the bottom of the screen, that quickly indicate the options available at that point in the process.

In the next panel the languages to be translated are displayed; the next panel offers a selection of 17 technical dictionaries that can be selected. The next panel allows for the creation of a new file or editing of a previously-created one. If, for example, a new file is to be created and the file name is typed in, the Enter key is pressed and a blank file appears on the screen. A press of F2 puts the user in the "Power-Typing" mode under XEDIT. Once a file has been created a press of F10 sends the file to be translated. The words

PLEASE WAIT WHILE I TRANSLATE

are displayed in the upper left-hand corner of the screen. The translated English text will appear on the screen in approximately 20-30 seconds, depending on the length of the file and activity on the main-frame.

If all words are translated without errors the translation will be displayed with the message

ALL WORDS WERE TRANSLATED.

If there are untranslated words from the original file the message

NOT-FOUND WORDS EXIST

will appear on the screen. To find these words, the user presses F9 and the not-found words are highlighted, in sequence, in the original input file by a row of asterisks above and below the not-found word. Once a correction has been made, a re-press of F9 brings up the next not-found word, and so on through the file. After all corrections have been made the user is told that no more not-found words exist, and to press ENTER to retranslate the file. The corrected file is then re-translated, in approximately the same length of time.

This error-correction process is a unique and widely-used feature of the menu. Both the input file and the translation file are permanently retained on disk and can be printed out on local printers.

Use of Interactive MT

As stated previously, the primary use of interactive MT is to provide rapid translations of short items for information scanning. The system is used by analysts and analyst assistants in a secure computer environment mode. Statistics have been kept on usage since the system became operational. The average number of accesses per month over a two-year period are:

Russian	185 accesses, with peak monthly use of approximately 625 accesses
German	27 accesses per month
French	25 accesses per month.

In addition to the primary use of the interactive system — a quick translation tool providing raw MT for information scanning by FTD analysts — the system has several peripheral uses, which will be briefly discussed.

1. The system is being made available to select Department of Defense components, contractors who support FTD research activities, and other organizations involved in the analysis and assessment of worldwide technological developments. These remote users can gain access to the translation software on FTD's mainframe using a modem and telephone lines. This application is relatively new, and no data currently exists on the extent of usage.

2. The interactive MT system is used as a spelling checker for correcting typos in large files to be translated by the batch MVS MT system. The rationale is very simple: the fewer mistyped words in a given text file, the faster the not-founding procedure and the more accurate the parse. Although Systran MT systems can tolerate a percentage of not-founds per page and the system automatically analyzes the function of a not-found word or typo, the greater the accuracy of the input text the better the MT results. We are now using the correction feature of the interactive system to clean up typing mistakes in all material that is to be sent through the batch MT system and then edited by translators. For example, a local external contractor who keys in Cyrillic text for FTD has dial-up access to FTD's mainframe. Once the text has been input it is shipped to FTD's mainframe and translated via the interactive system. The correction feature then highlights the typos, they are corrected, and the cleaned-up file remains on disk to be accessed and later processed by the batch system. It obviously takes more time to translate longer files through the interactive system. A file for batch processing can be from five to 20 pages long, and hence the clock time for running the file via the interactive system may be 15 minutes. But this is merely computer time. The gains have been significant in productivity by using the correction feature to produce files free of all typing errors.

3. The interactive MT system under CMS has been downloaded to run on various configurations of stand-alone IBM personal computers. To date the systems have successfully run on an IBM AT/370, an IBM AT/370 with an A74 processor box, and the new PS-2 7437 IBM workstation. In this application exactly the same menu is used as was described earlier. The potential use of stand-alone computers is evident: for users in remote locations without access to an IBM mainframe it is the perfect solution.

4. The interactive system, in the unclassified mode, is used within the Directorate of Translations in several ways. First of all, it is used as a quick diagnostic tool for developers of the system and lexicographers. But it is also used for very rapid turnaround of short documents that sometimes seem to get lost in the queue behind big books. Thus, one or two pages are input, a hard copy of the document is edited by a translator, the changes are entered into the English file, and the document is then printed and returned to the requester very quickly. Finally, the interactive system is used by quality-control personnel to translate omissions detected in larger translations when a large document is being quality-controlled. Use of the interactive system in the unclassified mode is steadily increasing. Total accesses now average 274 per month, with peak accesses approaching 600 per month.

The Future

We are planning to develop two software additions to our MT systems in the very near future which we believe will greatly increase the effectiveness of MT and its attractiveness to and use by end users.

The first thing we plan on doing is to give the end user the capability of creating his own dictionary. The dictionary will come in two forms: (1) a customer-specific PC dictionary, and (2) a customer-specific dictionary with topical glossaries, also PC-based. The first user-controlled dictionary allows the user to supply his own terminology on a PC which will supplement the main Systran dictionary with meanings and grammar codes for not-found words and replacement meanings for existing Systran dictionary entries. Individual words and word expressions may be entered. The customer dictionary with topical glossaries is similar, but it allows the capability of creating and modifying 16 specialty dictionaries in addition to the usual Systran technical glossaries.

The customer-specific dictionary, in both forms, is an override dictionary to the Systran main dictionaries; it resides in a buffer and allows up to 5000 entries. The dictionary at no time is permanently merged into the main Systran dictionaries, and hence the integrity of the Systran dictionaries cannot be jeopardized. However, the user does have the ability to fully control the translation of certain classes of words and phrases, and his modification decisions – for good or bad – will be reflected in the translations his system produces. This is a powerful tool in the hands of end users that must be used with foresight and care.

The customer-specific dictionary features a simplified and scaled-down version of dictionary coding and has a user-friendly menu. Modification is limited to adding new words and contiguous-word expressions, deletion of words, and modification of English meanings of existing words and expressions. Complex semantics, the ability to set bits and bytes on, scanning, and if/or statements are not allowed. The menu asks for rudimentary information about the source word(s), including gender, part of speech, declension class, and animation. The menu also queries information on number and declension in the target language. The goal is rapid dictionary development and control by the user without burdening him with the complexities of linguistics.

The customer-specific dictionary offers several distinct advantages to the end user. First, it permits the user to control certain aspects of the MT process, allowing him to become actively involved. There is much data confirming that the reception of MT by translators or analysts is increased when the user feels that he has some control over the translation process, that his corrections can improve the system, and that he is not totally at the mercy of an impersonal black box. Moreover, the customer-specific dictionary is an efficient way of resolving multimeaning or translation preference disputes among a wide audience of users. Finally, if any particular user has classified terminology that cannot be entered into the general Systran dictionaries, he can retain these terms in his user-controlled customer-specific dictionary.

We are also developing a post-processor for the finished machine-translated English file, beginning with English translation produced by the Russian system. The idea is to improve the readability of the final English file by automatically manipulating it to remove instances of awkward or ungrammatical usage that make the translation difficult to read. In a limited sense it would be a "translation of the translation," but would specifically address certain classes of errors produced by machine translation. It would incorporate some of the features of what are called grammar checkers; however, the errors would be corrected automatically. Based on empirical generalizations made from a large corpus of raw machine output a limited rule-based parser would be developed for English and incorporated in a post-processor.

Initial considerations would be the use of articles in English, treatment of noncopular sentences, animate/inanimate pronoun resolution, and rules to resolve certain slashed entries (*vse* = all/entire). An example of the latter is that in English one can say "all the boys" but one cannot say "entire the boys," or one can say "the entire group" but not "all the group." Although these issues are dealt with by the current MT software, the coding and linguistic software has become so complex that it seems easier to deal with certain classes of readability problems from a new perspective – the machine-translated English file. The goal is to write a limited English parser that will produce results without investing a great deal of money in the development. We believe this goal is attainable.

Conclusions

Machine translation has undergone a long evolutionary development at FTD extending over more than 20 years. The Russian system, e.g., has consistently provided fast-turnaround, economical, and usable translations that have met user requirements. Specialty dictionaries that provide consistency of technical translations have been continuously developed and updated over the years. New language pairs have been added. The recent development of offering machine translation directly to the user-analyst has dramatically broadened the scope and acceptance of machine translation. The use of rapid raw-machine translation for obtaining the essential information content of short texts is becoming more and more important. This seems to be the solution to dealing with the information explosion we are now witnessing. Finally, giving the end user local control over certain types of dictionary development on his own PC will, we believe, foster greater interest and use of the interactive system for rapid information acquisition.

STATE OF THE ART IN WEST GERMANY

by

Thomas Schneider
Siemens AG
DI AP 341
D-8000 Munich 83

In Europe, research and development in the field of machine translation has been boosted by the EEC's EUROTRA program. However, operative systems are usually based on older technologies, as in SYSTRAN or LOGOS. In recent years, the METAL system designed by Siemens has proved its applicability. As it exemplies the state of the art it is described in detail. METAL is a modular system with recursive grammars and non-sequential processing. It contains hierarchically structured lexicon modules to facilitate subject-specific translation. The end-user is provided with powerful tools to update his own lexicon. METAL is integrated into a chain of automated processes from the acquisition of the source text to the production of a camera-ready version of the target text. User experiences show a marked productivity gain and a reduction of turn-around time.

In 1966, the ALPAC report ended American government funding for the development of operative machine translation systems, citing the various project failures and pointing out that all the millions of dollars of support had not been able to establish a single operative system. Unfortunately, the positive suggestion to invest more money into theoretical basic research for machine translation was overlooked and most American projects were cancelled. European researchers were a bit less affected by the ALPAC report since they had not received large amounts of government funding anyway and were not under the pressure to produce large operative systems. Research in Europe centered mainly in Grenoble where under the direction of Bernard Vauquois GETA was established, and at the University of Saarbr}cken which eventually received funding from the German government for the development of the SUSY system. As with most university projects, it was not to be expected that commercially viable and robust systems would be designed. Lack of long-term financial support and personnel turnover were one of the reasons, the lack of adequate hardware for such applications as well as an insufficient linguistic basis were another. In retrospect it is somewhat strange that multilingual Europe had not been more active in the field of machine translation. The first commercially available system was Systran, designed in the USA, followed by Logos, also designed in the USA. Both systems are available in Europe. Systran is offered in France via Minitel through Gachot S.A., and the Logos system is marketed as a software package on IBM mainframes. Other companies such as Weidner (also US based) did not survive the extremely high investment necessary to come up with a marketable and viable product.

Since then, the further integration of the European Community has sharply increased the need for operative machine translation. Concurrently, the field of Computational Linguistics has finally established itself at various universities, from Leuven to Manchester, Bergen to Nancy and Stuttgart, to name just a few. In other words, the base for linguistic work towards the elusive goal of high quality machine translation has been broadened considerably. The EUROTRA project sponsored by the European Community may not result in an operative system in the near future. It nevertheless has done a lot to promote research in the field of Computational Linguistics and machine translation in particular. Certainly, the European public has become more aware of the problems imposed by multilinguality. As all national languages of the member countries are considered equal, a vast amount of documents must be translated to and from several languages. Already the European Parliament is spending more than half of its budget on translation. Outside of public administration, industry is equally affected.

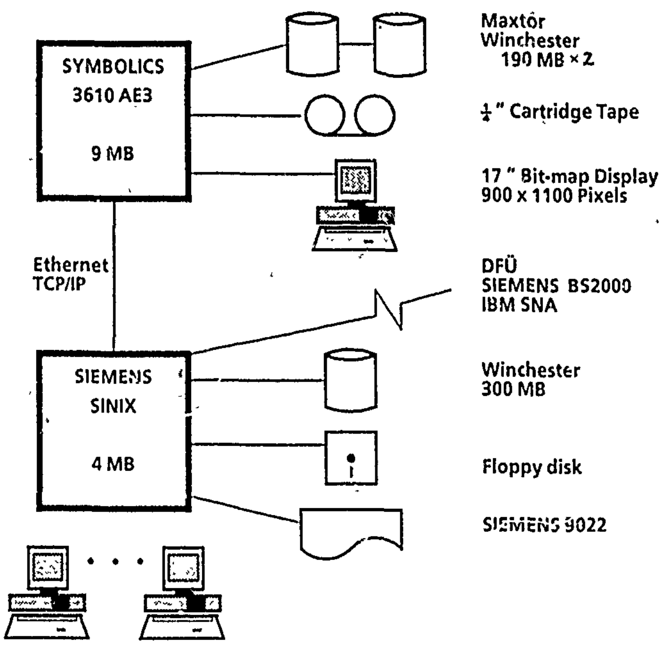
Costs for research and development are spiralling. At the same time, the life expectancy of a newly developed technology is decreasing. At the beginning of the century, a new technology could be expected to last for about five decades before being superseded by the next generation. By now a new technology may be obsolete in less than five years, and the innovation cycles are getting shorter still. In addition, our technology is changing, away from self-explanatory implements towards more and more complex products. The concrete and tangible objects of the past did not require extensive documentation since information about function and operation of the device could safely be assumed to be within the "world knowledge" of the user. However, with the advent of miniaturization of devices and a gradual shift towards abstract implementations of problem-solving tools and procedures like software, the user is no longer in a position to comprehend the workings of such a sophisticated system without explicit and detailed documentation.

This combination of ever shorter innovation cycles and an increasing amount of documentation per product leads to a veritable explosion of the volume of documentation in the industrial sector. The tremendous costs for research and development can only be recovered if larger markets are found, i.e. if export of a product can augment sales in the home market. This however necessitates the translation of the relevant documentation. Even within Europe there is no lingua franca which would be understood by all. Experiences show that among Europeans the presumed competence in a foreign language is usually overestimated. There are very few engineers who are able to understand a complex foreign-language description of a complex system with the degree of precision required for the error-free operation or even further development of such a system. The same holds true for the exchange of scientific research results. Unavailability of such results on account of language barriers can lead to the unnecessary duplication of effort or to costly errors.

Contrary to public belief, there is a noticeable shortage of technical translators which causes great concern in the industrial sector. To give an example: the complete documentation for a public switching system may amount to more than 100 000 pages. As on the average technical translators produce about 1000 pages per year, the task of translating this single set of documentation requires about 100 man years. Any company would be hard pressed to find sufficient qualified personnel, and even if twenty specialists could be found there would be a delay of five years between delivery of the physical product and its operation. Such delays can easily lead to the loss of markets. Therefore, besides the European Communities it is mainly in large businesses that the topic of machine translation has been addressed. Philips in the Netherlands is developing a prototype named Rosetta using isomorphic grammars, and the Dutch software firm BSO is working on a system named DLT which attempts translation via an interlingua based on Esperanto. Within Germany, the most headway has been made with the METAL system. As it exemplifies the state of the art it will be described in detail.

Siemens became involved in the area of machine translation in the late seventies. Experiments with commercially available systems proved less than successful so a decision was made to start a research and development project with the goal of building an operative machine translation system to increase the productivity of the in-house translators and reduce turn-around time. In 1978 Siemens entered into a cooperative agreement with the University of Texas at Austin. The Linguistics Research Center at UT was in the fortunate position of having been able to devote many years of research to contrastive and computational linguistics, without being forced to satisfy investors by marketing systems prematurely. The Center's work was conducted under the title of "METAL", and even though the present system bears no resemblance to the early versions the name has been retained. A first prototype was tested in 1979. The large program written in FORTRAN was loaded into the largest mainframe available at the university; all other users had to leave the system. In the experiment, one short sentence was to be translated from German to English, and only the pertinent lexicon was loaded. Still the system labored and labored until finally a translation appeared - after more than three hours ! On one hand, the experiment proved that the linguistic approach in METAL might work, on the other hand it showed quite clearly that an operative machine translation system needed to be designed and implemented in a different manner.

Hardware:
By now, the linguistic component of METAL is written in CommonLisp, the other functions such as the text processing component are written in C. The system is implemented on a hardware package consisting of several translator workstations and a dedicated LISP machine running as a server in the background. The hardware configuration looks as follows:


SYMBOLICS
3610 AE3
9 MB
Ethernet
TCP/IP
SIEMENS
SINIX
4 MB
Maxtor
Winchester
190 MB × 2
¼" Cartridge Tape
17" Bit-map Display
900 x 1100 Pixels
DFÜ
SIEMENS BS2000
IBM SNA
Winchester
300 MB
Floppy disk
SIEMENS 9022

Symbolics LISP machines are small enough for an office environment but very powerful. The translation throughput with METAL is about 200 pages per day. That is far more than a single translator could ever postedit. As the LISP machine is a single-user system it is linked via Ethernet to a multi-user translator workstation running under SINIX. From these terminals, translation jobs are started and all the tasks of deformatting and reformatting and postediting are handled. The translation process running in batch in the background is detached from other processing steps and does not interfere with any of the tasks at the translator's terminal. The SINIX system also provides the interface to other office systems, e.g. the Siemens or IBM office environment.* For reasons of lexicon integrity and uniformity of terminology, the functions of lexicon modification and structuring reside centrally on the LISP machine. This physically supports an organization where lexicon maintenance is performed centrally for an installation and ensures that responsibility for the lexicon remains with the terminologist in charge, without - possibly anonymous - interference.

---

* SINIX is the UNIX from Siemens.
UNIX is a registered trademark of AT&T.

System Structure

From the outset, METAL was built in a highly modular way so as to permit the inclusion of new elements or the modification of existing elements without major ill effect on the other components. There is a language-independent core system to which language-specific modules for analysis, transfer and synthesis are added. The analysis module of a given language is designed in such a way that it can be used as the basis for transfer to various target languages without any modification. This decreases development time and expense for new language pairs. Furthermore, the "open" system structure also makes METAL an adequate basis for future applications in semantic content analysis information retrieval or as a natural-language front-end for expert systems or data bases. Its first application, however, is machine translation.

Grammar

As there is at present no linguistic theory available that would describe even a single language unambiguously and completely a somewhat eclectic approach has to be chosen in the grammar. METAL employs a transfer system rather than an interlingua. It seemed that to define a meta-language incorporating all possible features of many languages would not only be an endless task but rather fruitless as well. Such a system would soon become unmanageable and perhaps collapse under its own weight. If on the other hand the intermediate meta-language were reduced to a manageable level of abstraction then too much surface information necessary for a faithful translation would be lost. Abstract formulae describing a text may be adequate for a rough paraphrase but not for translation with the aim of publishing the target document. Tests with several European languages have shown that at least between these related languages a transfer system is adequate. METAL uses basically phrase-structure rules which are augmented by tests on the constituents, their interaction and various other constraints. In contrast to other systems, the rules are recursively applied so that their number can be kept low. To illustrate the advantages of a recursive system let us take the following (simplified) sample rules:

rule 1 : S - NP VP
2 : NP - DET ADJ N
3 : ADJ - ADJ ADJ
4 : ADJ - ADV ADJ

Rule 1 says that a sentence may consist of a noun phrase (NP) and a verb phrase (VP), rule 2 that a noun phrase may consist of a determiner (DET), an adjective (ADJ) and a noun (N). Rules 3 and 4 on the other hand state that an adjective may consist of two adjectives, or of an adverb (ADV) and an adjective respectively (of course, all constraints and tests have been left out in our sample rules).
Now take the following sentences:

a. The old car runs.

Two rules, 1 and 2, would be necessary to interpret the surface structure as a sentence.

b. The very old car runs.

Here, rules 1, 2 and 4 would lead to a sentence analysis.

c. The rusty old car runs.

Rules 1,2 and 3 interpret the structure to be a sentence. According to rule 3, the two adjectives "rusty" and "old" are interpreted as one adjective for analysis in rule 2. If we continue to apply rules 3 and 4 to a given surface structure we can reach an interpretation of very complex structures, even of something admittedly contrived like:

d. The very rusty shabby slightly dented comfortable old car runs.

Imagine having to construct rules like NP - DET ADV ADJ ADJ ADV ADJ ADJ ADJ N to analyze a trivial sentence like this... A conventional machine translation system usually tries to account for every possible surface structure with a separate rule. This approach assumes (falsely) that a natural language is a finite system and that a sufficiently large set of individual syntax rules would eventually cover all cases. Aside from the fact that for free word order languages this is intrinsically impossible, managing tens of thousands of individual rules is very difficult. METAL at present uses no more than 600 grammar rules but is nevertheless able to deal with sentence structures it has never encountered before. On account of its recursive structure, the grammar does not need to state explicitly that a certain sequence of constituents is grammatically legal and may be interpreted as a unit. The grammar rules in METAL will generate legal structures from their base components. In other words, the METAL grammar is an "open" system whose coverage extends far beyond the explicitly stated rule content.

The grammar rules are indexed to make processing more efficient and also to allow the partial use of the grammar rules for e.g. "quick and dirty" translation for purposes of information gathering. The most commonly applied rules, e.g. those for word level morphology and for frequently occurring basic structures, are defined as the most basic level. Higher level rules deal with more complex or even ungrammatical structures. If a given surface structure can be analyzed using lower-level rules then the more complex and less likely rules are disregarded, which saves processing time. If no interpretation is possible with the lower level rules then incrementally higher levels of rules are added to the lower level rules, and again an interpretation is attempted. If for the purpose of a rough translation only the lower three levels of rules are invoked the translation result will not be as good but perhaps still quite adequate for some applications, and processing will be faster.

A second principle distinguishes METAL from older systems: linguistic parallel processing. Of course, it is impossible to translate at the word level. Not only may single words denote different concepts, as e.g. a "ball" may refer to a formal dance or a round object used in sports. Much more problematic is the fact that a word may reflect one of several different word classes, each with a different syntactic function in the sentence. The English word "back" for example can be:

noun:	His back hurts
adjective:	The back issue of Punch...
adverb:	Meanwhile, back at the ranch...
verb particle:	The boss paid him back
verb stem:	His colleagues back him up.

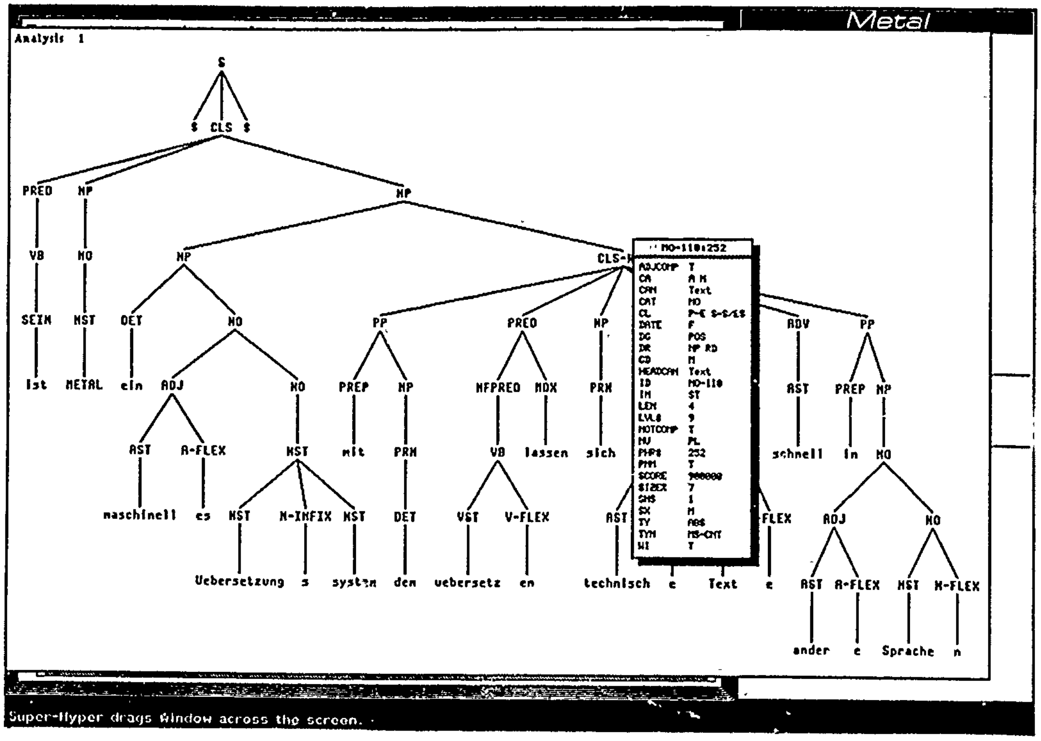
A decision about the function of "back" within the sentence cannot be made at the word level. However, even at the phrase level there would be problems. "Eating ice cream" may be considered a contiguous phrase, as in:

Eating ice cream can be pleasurable.

In a different context, however, the same surface string would not constitute a syntactic unit, as in:

Children eating ice cream can make a mess.

For a correct interpretation it is indispensable to analyze the complete sentence. This is esp[illegible]ial when dealing with free word order languages such as German, where one el[illegible]e verb may occur in a position quite distant from the other (separable prefixes, for example). In METAL, all possible interpretations of all elements in a sentence are written in a chart. The parser builds structures, utilizing the grammar rules and information contained in the lexicon. These structures are weighted based on probabilities and compared. Only when an interpretation spanning the whole sentence and accounting plausibly for all elements is reached, is the transfer to the target language attempted. In other words, no decision about the function of a sentence element is made until all other elements have been considered as well. This is computationally expensive but seems to be the only way to treat a natural language with all its ambiguities. If no interpretation spanning the whole sentence can be found the system invokes a fail-soft mechanism and delivers a translation of the individual phrases it had been able to interpret. In some language combinations the output may still be grammatically correct. In other cases, the posteditor has to correct the output. At the end of the analysis phase, the sentence is depicted as a tree structure. Behind each of the nodes is an extensive set of grammatical and lexical information:

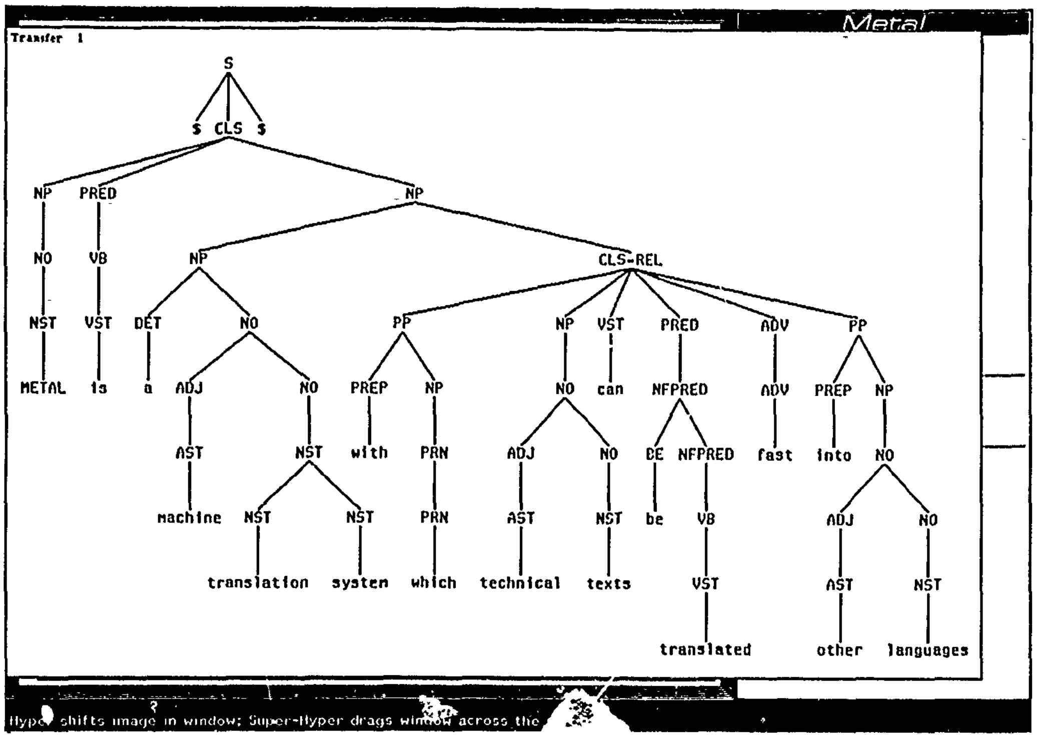

Metal
Analysis 1
S
$ CLS $
PRED NP NP
VB NO NP CLS-
SEIN NST DET NO PP PRED NP ADV PP
ist METAL ein ADJ NO PREP NP NFPRED MDX PRN AST PREP NP
AST A-FLEX NST mit PRN VB lassen sich schnell in NO
maschinell es NST N-INFIX NST DET VST V-FLEX AST FLEX ADJ NO
Uebersetzung s system den uebersetz en technisch e Text e AST A-FLEX NST N-FLEX
ander e Sprache n

NO-110:252

ADJCOMP	T
CA	A N
CAN	Text
CAT	NO
CL	P-E S-S/ES
DATE	F
DG	POS
DR	NP RD
GD	N
HEADCAN	Text
ID	NO-110
IN	ST
LEN	4
LVL#	9
NOTCOMP	T
NU	PL
PHP#	252
PNM	T
SCORE	900000
SIZE#	7
SNS	1
SX	M
TY	ABS
TYN	MS-CNT
WI	T

Super-Hyper drags Window across the screen.

In the transfer phase, this tree structure is transformed into a normalized tree structure appropriate to the target language:


Metal
Transfer 1
S
$ CLS $
NP PRED NP
NO VB NP CLS-REL
NST VST DET NO PP NP VST PRED ADV PP
METAL is a ADJ NO PREP NP NO can NFPRED ADV PREP NP
AST NST with PRN ADJ NO BE NFPRED fast into NO
machine NST NST PRN AST NST be VB ADJ NO
translation system which technical texts VST AST NST
translated other languages

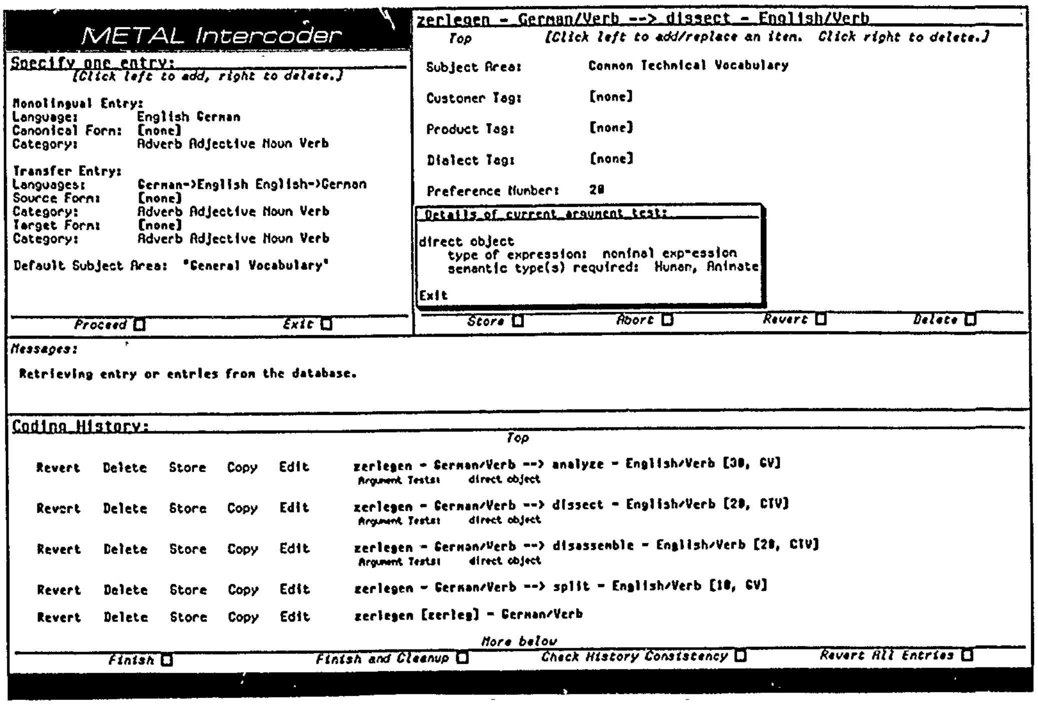
Hyper shifts image in window; Super-Hyper drags window across the

Out of this tree structure, the target sentence is generated.

Lexicon

No machine translation system can operate without an adequate lexicon. But the overall number of entries in a system dictionary is not a relevant criterion for a qualitative assessment or for a legitimate comparison of different systems. For one thing, the internal structure of an entry may differ.Perhaps all stems or even all tokens of a word are listed separately in the dictionary, or by contrast all forms may be subsumed under a single entry, with internal pointers to tables and rules so that full forms can be generated. METAL employs the latter structure.

Secondly, it makes a difference whether a system relies on one unidirectional dictionary, with a direct link between one source language word and one target language word, or whether multiple dictionaries are used. METAL operates on both monolingual lexicons and a transfer lexicon. The monolingual lexicons contain morphological, syntactic and semantic information needed for the analysis and/or generation of a language. The transfer lexicon provides a link from the source language to the target language, indicating under which conditions, in which contextual environment and in which subject field a source language entry should point to a specific target language entry. As an aexample, the German verb "zerlegen" would be translated into English as "analyze" if the direct object has the canonical form "Satz" (sentence). It would be transferred as "dissect" if the direct object has the semantic type 'human' or 'animate' but it would be translated as "disassemble" if the direct object is concrete.



Specify one entry:
[Click left to add, right to delete.]

Monolingual Entry:
Language: English German
Canonical Form: [none]
Category: Adverb Adjective Noun Verb

Transfer Entry:
Languages: German->English English->German
Source Form: [none]
Category: Adverb Adjective Noun Verb
Target Form: [none]
Category: Adverb Adjective Noun Verb

Default Subject Area: "General Vocabulary"

zerlegen - German/Verb --> dissect - English/Verb
[Click left to add/replace an item. Click right to delete.]

Subject Area: Common Technical Vocabulary
Customer Tag: [none]
Product Tag: [none]
Dialect Tag: [none]
Preference Number: 20

Details of current argument test:
direct object
type of expression: nominal expression
semantic type(s) required: Human, Animate

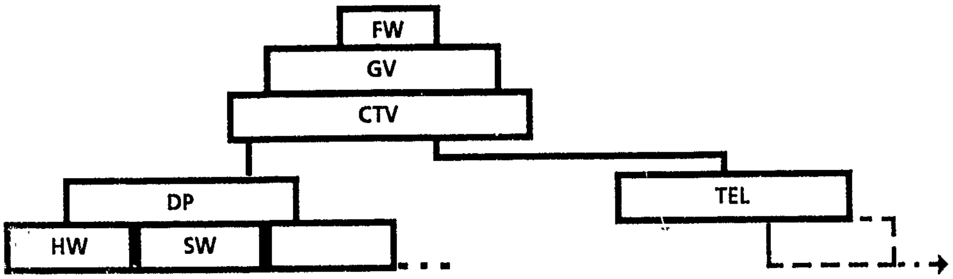
Messages:
Retrieving entry or entries from the database.

Coding History:
zerlegen - German/Verb --> analyze - English/Verb [30, GV]
Argument Tests: direct object
zerlegen - German/Verb --> dissect - English/Verb [20, CTV]
Argument Tests: direct object
zerlegen - German/Verb --> disassemble - English/Verb [20, CTV]
Argument Tests: direct object
zerlegen - German/Verb --> split - English/Verb [10, GV]
zerlegen [zerleg] - German/Verb

The advantages of such a lexicon structure as used in METAL are obvious. The extensive grammatical information contained in the monolingual dictionaries needs to be carried only once, even if many different entries in one of the languages correspond to the same entry in the other language. The transfers of the English verb "take" for example may fill several pages of a book. If each transfer entry were to contain all the morphological and syntactic information for "take" as well, the system dictionary would be inflated excessively. Not only would this waste storage space but it would also require superfluous coding efforts. Moreover, if monolingual and transfer lexicons are kept separate, the monolingual entries can be used in other language combinations without modifications.

Another aspect of a lexicon to be considered is the organization of its terminological content. In most European languages, the set of the most frequent 5000 words makes up approximately 90 % of any given text (on the average). Beyond this limited set, the point of diminishing returns is soon reached. Increasing an undifferentiated general lexicon to more than 100 000 words, for example, would not increase text coverage significantly. On the contrary, many unpleasant ambiguities would be introduced which can be avoided in a modular structure.

The METAL lexicon is organized as follows: There are modules for function words (FW) like prepositions, determiners and conjunctions, for general vocabulary (GV) and for common technical vocabulary (CTV) organized in a tiered hierarchy. From the next level down, each end-user can define and structure his own modules and tailor them to his specific application. For in-house applications in Siemens, there are for example modules like Data Processing (DP) with submodules Software (SW), Hardware (HW) etc. Furthermore, it is possible to define transfers on the basis of a specific customer, a specific product or a specific target country. Thus a text translated into British English will show "lorry" instead of "truck" for the USA, and a text intended for Spain will automatically have "ordenador" instead of the Colombian "computadora". The METAL lexicon structure can be visualized like this (simplified):



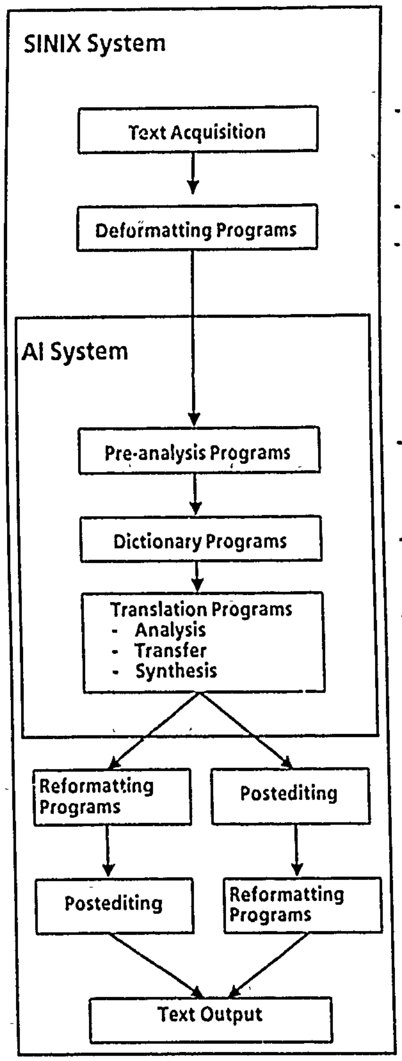
Before a translation run is started, the modules appropriate to the subject area of the text are defined. If the syntactic and semantic criteria for the selection of a lexicon entry are met and there are several candidates for transfer, then the one tagged for the subject area of the text or tagged for a hierarchically closer module is chosen. This assures that highest priority is assigned to subject-specific transfers.

The main source for the required terminology for new subject fields is TEAM, the multi-lingual terminology database operated by Siemens which at present contains approximately three million records in up to eight languages. An interface between TEAM and METAL facilitates the installation of new lexicon modules.

External users update their own lexicons with the aid of the so-called INTERCODER, an integrated expert system. It guesses at the morphological and syntactic behavior of new lexicon entries and proposes the necessary coding; the missing pieces of information are inferred from a set of rules and partial information already contained in the lexicon. The INTERCODER has proven its usefulness in reducing coding time by a factor of ten. While it is not recommended to alter function word entries (they are too closely linked to the grammar) a translator may code all other word classes including verbs. Even though the grammar rules are not accessible to an end-user, the transfer lexicon permits significant syntactic transformations. On top of being able to specify transfers on the basis of the instantiation of frames, the presence of arguments of a certain semantic type or a specific canonical form, the user can influence the target structure considerably. Source language active structures may be turned into impersonal constructions, roles of arguments can be changed, complements can be converted, elements can be added or deleted etc. Great care has been taken in the design of the user interface so as not to overburden a translator with linguistic detail.

Office Environment

An operative productive system needs to do more than simply translate individual sentences entered from the keyboard. Most of the texts which have to be translated quickly and are of great volume such as e.g. technical documentation are heavily formatted. In some texts more than half of the characters on a page may be non-translatable material, notably flow charts, diagrams, tables and various control characters for format and layout. It would be highly uneconomical to manually extract the text portions to be translated and afterwards manually re-input them. That would not only be expensive but it would also invite errors in the additional reformatting tasks. Therefore, METAL has been integrated into a chain of processes, from text acquisition via automatic deformatting and translation to automatic reformatting procedures. A translation run usually goes through the following steps:


SINIX System
Text Acquisition
Deformatting Programs
AI System
Pre-analysis Programs
Dictionary Programs
Translation Programs
- Analysis
- Transfer
- Synthesis
Reformatting Programs
Postediting
Postediting
Reformatting Programs
Text Output

- Text transfer via data link or input facilities (floppy disk, magnetic tape, page reader)
- Separation of language and format data
- Processing of special formats (diagrams, tables)
- Determination of words to be added to system dictionary
- The INTERCODER - an interactive expert system
- Translation
- Merging of language and format data
- Revision of translation
- Word processing system
- Printer output
- Typesetting

A text is usually received in machine-readable form, by file transfer, floppy system check the pages for tables, graphs etc and mark them. They identify the text portions to be translated and generate a mask of the page. The individual translation units, usually sentences but in the case of headlines or table entries also single words or phrases, are automatically recognized, numbered consecutively and extracted from the page mask. They are written into a text file and transferred to the LISP machine for translation. After translation, the file containing the target language text units is returned to the SINIX system for post-editing. Here, the translators can choose whether they want to postedit an interlinear version which groups single source language/target language units sentence by sentence, or work on two windows with source and target text, or whether they prefer a target language output that has already been reformatted. In the former cases, the posteditors would start the reformatting program after having made their corrections. At the end, the target language text is available with all the formatting information and with the same layout as the original.

Before a text is translated, it is advisable to run a comparison of text and system lexicon. As linguistic processing is based not only on grammar rules but also on information contained in the lexical entries, sentences in which several words are unknown to the system are difficult to analyze, and the translation is likely to be inferior. Therefore missing words should be added to the lexicon. In METAL, the comparison of lexicon and text produces several files. One is a list of unknown words, each listed with its location and context so that transfers are more easily found. This list will actually also show faulty orthography so that the program can be used as a spelling checker as well. The second output is a list of compound words which were not found in the system lexicon but for which a translation is proposed on the basis of the individual components. Here the translator is called on to make sure that the proposed translations are appropriate to the subject area. The third output is a text-based glossary, listing source term and proposed translation. This may be used to review subject area adequacy of the lexical entries. It is also useful if, in a large document, one portion is to be translated by the machine translation system and the initial pages are written in a style which makes them unsuitable for machine translation. In such a case, the human translators can be given a glossary of exactly the terms contained in the pages to be translated so that they don't have to wade through mounds of subject-area listings. This will ensure that the same terminology will be used throughout the whole document.

Quality and User Experiences

The state of the art in computational linguistics does not permit the perfect translation of random texts. Therefore, if a text is translated not simply for the purpose of getting a rough idea of the content but with the aim of publication, postediting by a human translator will remain a necessity. Even if a system is tuned for specific subject areas there are still sufficient problems in linguistic analysis, especially if the meaning to be conveyed is hidden "between the lines". One should not attempt to measure the "correctness" of machine translation in percentage points. Just as with human translation, there is not necessarily a single solution. The quality of a translation does not hinge on the quality of the translation system alone but is equally dependent on the quality of the source text. Inputting garbage will not produce poetry. One also needs to consider the intended purpose of the text, expectations of the readers and even the stylistic preferences of the post-editor.

The quality of a machine translation system can only be judged in regard to the questions if translators working with the system have been able to increase their productivity and decrease turn-around time. One prerequisite of course is the willingness of translators to use the system in their daily work, and that presupposes not only a fairly high level of translation quality but ease of operation as well.

Machine translation is a recently evolved technology and is as such vulnerable in its status. A new technology can easily be proven inadequate or even useless if the intended recipient refuses to accept such a system or insists on applying it in unsuitable ways.

Therefore the introduction of a machine translation system into an existing organization, be it a large industrial company or a translation bureau, requires several steps. First of all, end-users must have a clear picture of what can be expected from an MT system and what is beyond the scope of today's technology. Inappropriate use will only lead to frustration.

Once the conditions for the installation of a system have been assessed, i.e. translation volume, suitable types of text, hardware environment, and a positive decision has been reached, the organizational setup needs to be discussed. From which sources does the translator receive the original texts? Is there a possibility to influence the style of the original, to impose certain guidelines in regard to complexity of verbal expressions? And can the customers be persuaded to use standardized formatting and layout routines so that the tasks of deformatting and reformatting can be simplified?

Translators using machine translation systems need an introductory training. It should focus on a general introduction to the system's structure and the tools it provides. Equally important is a first training in the different work techniques that such a system requires. Provided that the reader of a target document is not concerned with intricacies of style, the post-editing phase of a machine output can focus on changing this output to an acceptable version with the least effort. Certainly, a given version could be rewritten in various ways, sometimes with a gain in quality but sometimes also with simply an idiosyncratic change of style without improvement of quality.

Postediting machine output is different from revising a "human" translation. While the machine will make "severe" errors in syntax, e.g. in prepositional phrase attachment, or semantics in ambiguous structures, a human translator will make fewer but random and less predictable errors. Usually, it takes a translator several weeks of practical work with an MT system to be able to anticipate the common errors perpetrated by the system and look for them. Experiences with more than a dozen METAL installations have been quite positive and can be summarized as follows:

Translators as well as upper management have to understand that a machine translation system is not a substitute for a highly qualified translator but no more and no less than a powerful tool.

For the use of METAL, an initial training period of one week has been sufficient. A second week of training after a few months answers questions which have arisen during the actual productive application. After that, consultation on a case by case basis seems adequate.

During the first few months of operation, the translators' productivity will actually decrease. There is the initial overhead of bringing the lexicon up to a level where it covers most of the specific texts to be handled. Also, translators have to get used to the different work technique and acquire skills in lexicon building and system administration.

After this initial learning phase, which may vary from a few months to more than a year, users have reported considerable gains in productivity and a decrease in turn-around time. It appears that under favorable conditions a productivity gain of a factor 2 to 3 is a realistic goal. In addition to the benefits derived from increased productivity, the consistency of terminology throughout all documents has been viewed as a qualitative improvement of the target text which could not have been achieved with "human" translation.

METAL is now available as a product. Development will continue to integrate additional language pairs and to streamline the interface to various office environments. Further research will focus on add-on semantic components and linking METAL to data bases, expert systems and teaching/learning systems. Even if the state of the art does not permit the ideal solution in the area of natural language processing it seems that systems such as METAL can contribute decisively to an improvement of multilingual communication.

Synthèse des solutions proposées aux utilisateurs

Albert YANEZ
Conseiller du Directeur
Centre de Documentation de l'Armement
26, Bd. Victor,
75996, Paris-Armées.
France.

Résumé:

L'utilisateur potentiel de la TAO, c'est-à-dire chacun d'entre nous, et plus encore celui dont le métier est de traduire, doit pouvoir connaître, atteindre et utiliser la solution la plus adaptée à son cas particulier afin de réduire l'obstacle linguistique qu'il rencontre à son niveau pour communiquer au-delà du ghetto de sa propre langue.

On trouvera donc dans cette synthèse une énumération de solutions constituant des recours possibles et allant du dictionnaire ou de la banque de données terminologiques jusqu'à des logiciels de gestion lexicographique ou d'analyse de texte ou d'aide à la traduction, étant entendu que dans tous les cas l'utilisateur aura à apporter un concours. Les résultats seront tantôt directement utilisables, tantôt frustres, de sorte que, selon le cas, on pourra s'en contenter ou bien on aura à faire appel à une autre assistance, celle d'un traducteur ou d'un réviseur humain.

En attendant le système parfait et idéal de demain, qui pour longtemps encore restera en laboratoire, il n'existe pas d'autre issue que d'établir un pont, une coopération étroite entre concepteurs ou vendeurs de systèmes d'une part et utilisateurs d'autre part. C'est de cette coopération qu'émergeront les solutions. Elles ne sont pas offertes: il faut les construire.

C'est pourquoi, au-delà des "produits" (outils ou systèmes) présents sur le marché de l'industrie de la langue et de la TAO en particulier, on met l'accent sur des aspects socio-politico-économiques qui sont loin d'être négligeables si l'on veut atteindre de vraies solutions, c'est-à-dire une situation où l'introduction de ces outils et de la TAO devient facteur de productivité, d'ouverture et de progrès. Or cette situation a déjà été atteinte par certains utilisateurs qui ont su faire un pas pour ouvrir la voie dans laquelle d'autres peuvent aussi s'engager, seuls ou en concertation.

ooo0ooo

1. Examen des solutions éventuelles.

Solution N° 1: pousser en avant sa propre langue. Cette solution n' apporte pas la réponse. D'abord parce qu'elle conduit à terme à la dégradation lente mais certaine de cette langue, qui ne s'appuie plus sur l'identité culturelle d'un même peuple , et se trouve utilisée par des partenaires en situation d'inégalité, ce qui engendre malentendus et frustrations. C'est néanmoins une manière de renforcer le lien entre peuples amis ou cousins.
Côté anglais, on a en mémoire les remarques récentes du prince de Galles qui déplore la dérive de la langue anglaise, un certain laxisme qui se répand dans son usage. Côté francophonie on ressent la nécessité d'une certaine vigilance (Comité International de la Langue Française(CILF)...

Solution N°2: Rechercher un interlangage comme la langue artificielle préconisée par l'International Auxiliary Language Association (éléments linguistiques communs aux langues romanes et à l'anglais) ou une lingua franca universelle comme l'espéranto du Dr. Zamenhof. Solutions utopiques, rappelées néanmoins ici pour mémoire.

Solution N° 3: Encourager et faciliter l'etude des langues, rendre les gens polyglottes. C'est sans aucun doute une bonne chose mais en soi insuffisante. On croit connaître une langue étrangère alors qu'on n'en connait que les rudiments, et le problème demeure quand on est placé dans un cadre de communication professionnelle dans une langue non vernaculaire. Il n'empêche que nombre de pays, dont la France, pourraient utilement aller plus loin dans l'enseignement des langues, notamment chez les ingénieurs. Il se trouve qu'aujourd'hui de nombreux outils existent pour faciliter cet enseignement dès le stade de la petite école, par exemple CD TEL, ou l'enseignement assisté par ordinateur s'appuyant sur le minitel et le disque compact.

Solution N° 4: Faciliter l'accès aux dictionnaires électroniques et banques de données et autres outils terminologiques.

La plupart des dictionnaires mono ou multilingues de bon renom et d'usage courant sont numérisés et accessibles en ligne ou sur microordinateur à partir d'un disque compact, par exemple Collins-on-line, distribué par Softissimo (France), Robert, Hachette, New Oxford English Dictionary (NOED), Aérospatiale. Il en est de même des banques de données terminologiques:

- Eurodicautom, sur serveur Echo, contenant des centaines de milliers de termes et phrases de contexte et des dizaines de milliers d'abréviations,

-Normaterm, contenant 100000 termes français et anglais extraits des normes françaises et internationales et des textes règlementaires,

-Termdok, sur disque compact, donnant accès à 225000 termes avec définitions, en huit langues, et regroupant sept banques de données terminologiques,

-Termium, conçu d'abord pour vérifier et normaliser la terminologie dans les deux langues du Canada, mais également comme système d'aide aux traducteurs,

-Termnet, réseau international pour la terminologie, qui produit et diffuse des publications ou des produits et services dans le secteur de la terminologie, à l'échelle internationale.

-TDB (terminology data bank) intégrée dans un système d'aide aux traducteurs (Carnegie Mellon University).

La liste n'est probablement pas exhaustive mais il s'agit d'outils avec lesquels beaucoup de traducteurs sont déjà familiarisés, qu'ils aient ou non par ailleurs recours à la TAO.

L'annexe 1 indique le nom, le contenu et un point de contact pour chacune d'elles.

Solution N° 5: Utiliser des logiciels disponibles dans l'environnement de la traduction et de la terminologie: création, gestion et consultation de terminologie. J'en citerai quelques uns mais là encore la liste sera loin d'être exhaustive:

-Aquila, avec utilisation possible sur micro dans un éventail de 15 langues, distribué par La Maison du Dictionnaire (France),

-BDTAO, pour la gestion des bases de données lexicales, distribué par B'Vital(France).

-Alexis, de GSI-ERLI, permettant de naviguer entre des termes et des concepts,

- Ink Textools et Term Tracer, distribués par Ink Languages (France),

-Lexm 2, poste de travail pour lexicographe, distribué par SEI (France),

-Microcezeau qui permet notamment de fusionner des banques de données entre elles et d'échanger des données avec Eurodicautom dans de nombreuses langues, distribué par Terminformatique (France),

-Termex pour la création et la gestion de dictionnaires électroniques avec un programme complémentaire Glosnost, conçu aux Etats-Unis , distribué par Eurolux Computers (Luxembourg).

-Phenix: à chaque terme correspond une fiche terminologique reprenant les données contextuelles ainsi que des précisions grammaticales et lexicales (français, anglais, allemand, espagnol, italien), distribué par SITE (France)

-Thesaurus multilingue électronique distribué par Lexitech Utrecht (Pays-Bas).

Solution N° 6: Autres outils périphériques du traducteur:

-Système bilingue, qui permet l'usage multilingue des microordinateurs: reconfiguration du clavier, impression des caractères nationaux... distribué par microcoque Inc.(Canada) avec...

-EGA-Font, pour l'affichage de caractères nationaux ou de graphes scientifiques ne figurant pas dans les caractères de base.

-Ted, environnement de traitement de texte spécialisé pour la traduction: mise en fenêtre du texte source, du texte cible et de la traduction,; etc...distribué par Ink Langages et

-Textcount, logiciel de facturation automatique pour traducteurs, avec comptage des mots ou des lignes...distribué par Eurolux Computers (Luxembourg).

Solution N° 7: Avec la multiplication des banques de données, des réseaux télématiques et des passerelles qui les rendent aisément accessibles, on observe que, dans l'environnement de la recherche documentaire, il y a place pour des solutions facilitant l'identification de l'information utile dans un contexte multilingue. Il s'agira par exemple tout simplement d'indexation multilingue (fichier Pascal du CNRS, ou PERINORM de l'Afnor) mais surtout de l'intégration dans les logiciels de recherche documentaire de modules analyseurs de langues, s'appuyant sur des bases de connaissances multilingues et permettant en quelque sorte l'indexation automatique du texte entré et sa recherche dans l'une des langues acceptées par le système. C'est le cas de DARWIN, conçu et distribué par la société CORA (France). On peut ainsi, sans connaître la langue du corpus documentaire, interroger ce corpus dans une autre langue et obtenir des résultats plus précis et pertinents que ce que permet une recherche de type booléen à partir d'une indexation s'appuyant sur un thésaurus multilingue et des opérateurs de proximité.

Peuvent être compris dans ce type de solution les logiciels de routage de messages qui opèrent par détection des concepts correspondant à des destinataires et utilisant eux aussi un analyseur comparable à celui que l'on retrouve dans les systèmes de TAO.

Solution N° 8: C'est la possibilité offerte à un traducteur indépendant ou à une entreprise de mettre en place en interne, et dans les limites de son domaine d'activité, un système d'aide à la traduction sur microordinateur, c'est-à-dire en utilisant un investissement qui aura déjà été fait par ailleurs, par exemple pour le traitement de texte, ou l'interrogation de banques de données terminologiques, ou d'autres applications telles que celles qui sont citées ci-dessus. Il s'agit de systèmes tels que Alps ou Wiedner ou Bravice, que le producteur fournit avec un dictionnaire général et éventuellement des dictionnaires spécialisés, et une formation à l'utilisation du logiciel. Bien entendu, si le texte entré n'est pas déjà sur support magnétique ou s'il ne peut arriver par transfert de fichier en ligne (téléchargement), on sera conduit à adjoindre au poste de travail un lecteur optique assurant la reconnaissance de caractères, du genre Inovatic, en prenant soin de s'assurer qu'on bénéficiera ensuite systématiquement des progrès réalisés sur le logiciel, car les choses vont vite dans ce domaine et l'on risque d'avoir à brève échéance une installation obsolète.(Voir en annexe 2 les principaux logiciels de reconnaissance de caractères disponibles sur le marché français)

On devra aussi savoir qu'une réaction permanente avec le système devra être assurée afin de compléter les dictionnaires au fur et à mesure que leurs lacunes seront constatées. On peut alors associer au logiciel TAO un logiciel de traitement ou de gestion ou de navigation dans une base syntaxico-lexicale si l'on veut perfectionner le système et ne pas s'en tenir à des traitements trop sommaires.

Un préalable indispensable sera aussi de se renseigner auprès du fournisseur du système, et aussi auprès d'autres utilisateurs de ce système, qui ont pu développer eux mêmes des outils analogues et qui seraient intéressés par une coopération pour réduire leurs propres coûts.

Enfin il faut savoir aussi que des documents très courts (quelques pages), non numérisés préalablement, conduisent à un ensemble de manipulations qui ont pour résultat d'abaisser la productivité et de rendre à terme contestable le recours à la TAO qui au contraire se justifie pleinement si l'on aménage le poste de travail en veillant à son ergonomie.

Solution N° 9: C'est celle qui peut s'appliquer à l'information générée par l'entreprise; elle va de la publicité à la documentation technique accompagnant les produits et services. C'est l'ensemble des flux d'information sortants. Cette information a ceci de particulier qu'elle couvre un secteur délimité, bien maîtrisé, où l'on est orfèvre ou expert et donc tout à fait capable de définir et de contrôler le sous-ensemble lexical nécessaire et suffisant, et éventuellement même constituer un sous-ensemble syntaxique, en relation avec un guide de style ou de rédaction. Il est probable qu'il existe déjà dans l'entreprise toute une chaîne d'édition passant par la numérisation et un ensemble de contrôles. Il est possible aussi que l'entreprise ait à protéger une partie de sa production documentaire et que des problèmes de confidentialité existent bien que l'on ait à traduire, dans le cadre par exemple d'accords de coopération internationale. Dans un tel cadre de besoin, la TAO devra être un outil interne capable de s'intégrer facilement dans un processus d'édition et devra pouvoir accepter des dictionnaires constitués pour des besoins internes. Cette capacité à s'intégrer pourra alors constituer un critère de choix important. Au-delà des petits systèmes tels que Wiedner qui sont insuffisants vis-à -vis de gros volumes, on pourra donc envisager d'implanter en interne un système de traduction plus puissant. Un tel projet ne peut être économiquement viable que s'il est étudié en concertation par un groupement d'utilisateurs, et à cet égard l'exemple du CIGREF (Club Informatique des Grandes Entreprises Françaises) est extrêmement interessant parce qu'il apporte, vis-à-vis du concepteur ou du distributeur de système, un poids suffisant pour obtenir les aménagements souhaitables, et définir en commun une doctrine de développement et d'utilisation de l'outil TAO intéressant la collectivité.

Solution N° 10: A l'inverse du cas qui précède, il existe un autre type de besoin qui concerne les flux d'information entrants. En prticulier l'interrogation des banques de données textuelles qui aujourd'hui sont surtout de type signalétique mais qui de plus en plus offrent un accès au texte intégral. L'utilisateur a besoin de pouvoir faire un balayage rapide de ce contenu textuel pour identifier, à partir d'une recherche en ligne ou d'une diffusion sélective de l'information en ligne, établie sur son "profil" d'activité des informations qui seront tantôt dans sa langue, tantôt dans diverses langues étrangères. Dans une première étape, il est placé en face de résumés, généralement de langue anglaise, provenant d'un ou plusieurs serveurs d'information. On peut alors intégrer au niveau d'une passerelle ou "gateway" une possibilité d'accès en ligne à un serveur de traduction assistée pour présenter à cet utilisateur, lui-même expert dans le domaine considéré, une traduction brute dont il pourra généralement se contenter, en attendant de pouvoir faire traduire avec plus de soin le document primaire qu'il aura ainsi pu identifier plus facilement que si la base consultée est dans une langue qu'il ne connait que très mal. Le CEDOCAR a entrepris, sur ce type de besoin et de solution, des essais avec Systran, où l'on traduit en essayant de regrouper aussi bien les volumes que les transactions. On peut bien entendu imaginer que l'ensemble de la banque de données soit mis en traduction par son producteur: c'est là le sujet d'une étude économique qui reste à faire.

Solution N° 11 Il est arrivé qu'une entreprise ne trouve pas de système de traduction correspondant à son besoin, en l'espèce le besoin de produire une banque de données bibliographique en plusieurs langues et de pouvoir l'interroger dans ces différentes langues. L'entreprise peut alors créer elle-même son propre système de traduction automatique, puisqu'il ne s'agit plus ici de TAO. C'est un acte de foi mais il n'est pas interdit de penser qu'un tel système puisse intéresser d'autres secteurs industriels où l'on travaille aussi dans un contexte multilingue, à la production d'une banque de données en commun, auquel cas seule la base lexicale serait à revoir.Le cas de l'Institut Textile de France auteur de TITUS, opérationnel depuis plusieurs années malgré les contraintes imposées aux rédacteurs, mérite une pause.

TITUS vogue vers une version V qui sera incessamment en service, où ces contraintes seront très faibles et tout à fait acceptables.

Choix d'un système de TAO.

Les critères qui entrent en ligne de compte, avec une pondération qui reste à déterminer, sont les suivants:

-niveau d'intelligibilité du résultat brut, en général, et dans le secteur considéré, si l'on est dans une activité sectorielle.

-couples de langues acceptés par le système, directement ou par une autre langue interposée.

-vitesse de traitement.

-volumes à traduire (dans la situation présente d'une part, dans l'hypothèse de l'utilisation de la TAO d'autre part) ce point étant bien entendu en relation directe avec le critère qui précède.

-volume, qualité , accessibilité et facilité de mise à jour et de correction des outils linguistiques (dictionnaires, thésaurus, bases de connaissances) et possibilité de navigation entre ces outils en foction du contexte (intelligence artificielle).

-possibilité d'utiliser, en sous-produit de la mise à jour des dictionnaires, un produit de paramétrage sur support magnétique réutilisable éventuellement dans le cadre de l'évolution d'autres systèmes de TAO, même concurrents.

-facilité d'intégration dans la chaîne de traitement documentaire.

-compatibilité avec l'équipement informatique existant et le réseau de transmission de données.

-autres utilisateurs du système, et éventualité d'une association avec eux.

-aspects confidentialité.

-nombre et niveau de qualification des personnels associés au fonctionnement du système, y compris réviseurs. Coût de la formation nécessaire dans chaque qualification.

-ergonomie du système et niveau d'acceptabilité par les traducteurs et par l'utilisateur final (s'il n'y a pas phase de révision).

-possibilités d'apprentissage et de perfectionnement du système. (organisation de l'enrichissement des dictionnaires, niveau de complexité et de coût de cet enrichissement).

-possibilités de prise en compte des corrections syntaxiques.

-risques d'interférence et de perturbations entre plusieurs utilisateurs du système.

-prix d'acquisition ou d'utilisation du système, et valorisation des apports de l'utilisateur, par exemple dans le cadre de la constitution et de l'évolution de dictionnaires susceptibles d'être utilisés par d'autres (redevances ou ristournes).

-évaluation comparative du coût TAO et du coût traduction sans TAO pour 100 mots et des avantages et inconvénients de chaque solution (volumes, délais) projetés sur quelques années.

On trouvera ci-joint un tableau (annexe 4) qui présente les principales caractéristiques de quelques systèmes opérationnels. Ce panorama des solutions possibles n'est probablement pas exhaustif. Pour aller plus loin on pourra avoir recours au point de contact qui est généralement indiqué. On ne porte pas ici de jugement de valeur sur ces solutions, d'autant plus que l'efficacité dépend toujours de la nature de l'application et d'un environnement qui peut être très différent d'une application à l'autre.

Peut-être est-il utile également de fournir une indication sur les tarifs de rémunération pratiqués vis-à-vis des traducteurs. Les chiffres qui apparaissent dans l'Annexe 3 avaient été fournis en 1982 par Loll Rolling, de la CCE Luxembourg et auraient donc à être réactualisés. On trouvera aussi dans l'annexe 3 des éléments de comparaison de coût entre traduction machine et traduction humaine.

Conclusion:

Il existe aujourd'hui un certain nombre de possibilités d'améliorer la productivité en matière de traduction, depuis le simple recours à des dictionnaires électroniques ou autres outils linguistiques installés localement ou accessibles en ligne, jusqu'à la TAO proprement dite, en passant par des solutions intermédiaire comme l'indexation multilingue ou les analyseurs de texte s'appuyant sur des bases de connaissances multilingues ...Le souci de mise en commun des traductions effectuées (World Translation Index) et même des traductions entreprises va aussi dans le sens de l'amélioration de la productivité.

Il faut éviter désormais l'obstacle majeur et le surcoût qu'a été la saisie du texte, ce qui signifie qu'il faut générer le texte sur support numérisé, ce qui aujourd'hui fort heureusement tend à se généraliser. C'est bien entendu lorsque l'on est en présence de texte déjà numérisé que des gains substantiels peuvent être escomptés. Ceci signifie qu'il faut se tourner vers l'édition électronique et ne pas continuer à s'en tenir trop longtemps encore au seul support papier.

Il faut viser à intégrer la TAO dans la chaîne de traitement documentaire et la placer de préférence dans le service d'information, où l'environnement est le plus favorable, ce qui permettra des économies dans l'investissement.

Il faut que les traducteurs et interprètes soient plus étroitement associés non seulement comme utilisateurs mais aussi pour apporter leur compétence en matière d'enrichissement des contenus sémantique et syntaxique.

Il faut veiller tout particulièrement à l'ergonomie des systèmes installés, pour obtenir un confort d'utilisation suffisant.

Il faut par ailleurs se préparer à l'industrialisation de la langue en considérant que la TAO n'est qu'une application d'un effort plus général, intéressant d'autres secteurs de la communication. Cela étant, la TAO va pouvoir bénéficier de tous les progrès réalisés à d'autres fins dans le domaine de l'analyse et du traitement de la langue.

Il faut enfin procéder en Europe et aux Etats-Unis à une réévaluation des enjeux et du marché potentiel pour ne pas laisser le champ libre dans ce secteur au Japon qui a aujourdhui une appréciation toute différente de ce marché, de l'intérêt de la TAO et de la nécéssité de la faire progresser.

Il faut aussi faire appel aux compétences des traducteurs et interprètes, aussi bien au niveau du développement des systèmes qu'au niveau de leur utilisation.

L'enjeu a probablement été jusqu'ici sous-estimé aux Etats-Unis et en Europe. Tout indique qu'au contraire au Japon on investit beaucoup plus dans ce secteur non seulement parce que l'on espère exporter des systèmes de TAO mais surtout parce qu'on voit dans la TAO la seule manière de réduire très sensiblement l'obstacle de la langue, tant pour s'informer que pour se faire connaître.

En attendant que des systèmes évolués tels qu'EUROTRA voient le jour, probablement dans cinq ou six ans au mieux, il importe de satisfaire la demande de traduction, aujourd'hui de plus en plus pressante. Chaque utilisateur ne peut pas à lui tout seul faire tout l'effort nécessaire pour enrichir les systèmes existants. Il convient donc d'opérer des regroupements d'utilisateurs pour sélectionner le ou les systèmes qui méritent d'être enrichis.Dans la mesure où il faudra encore entrer des dictionnaires, ces utilisateurs peuvent faire ensemble le choix de ces dictionnaires, en privilégiant là encore ceux qui existent déjà sur support magnétique, et en recherchant une méthodologie permettant un paramétrage indépendant du système de traduction, de façon que le résultat de cet investissement soit utilisable pour d'autres systèmes éventuellement.

ooOoo

Bibliographie:

Répertoire des produits et services de traitement automatique de la langue française. Observatoire des industries de la langue. Editions Daicadif; ISBN: 2-90603606-4 (31 Janvier 1989)

Traduction assistée par ordinateur. Observatoire des industries de la langue. Actes du séminaire international, Paris Mars 1988 et dossiers complémentaires. Editions Daicadif; ISBN2-906036-05-6.

----oOOO----

Annexe 1:
Banques de données terminologiques

Nom	Contenu
Eurodicautom	> 370000 termes et expressions contextuelles > 90000 abbréviations mise à jour mensuelle (2000 entrées) manuel gratuit en anglais et en français
Serveur Echo, 15, Av. de la Faiencerie, L 1510 Luxembourg Tel +35220764	
Normaterm	100000 termes français et anglais extraits de normes françaises et internationales et de textes règlementaires. Accès par le français ou l'anglais (définitions, synonymes, termes génériques et specifiques, renvois, indication de sources...) messagerie associée - sur 3617 code normaterm.
AFNOR, Tour Europe, Cedex 7 92080 Paris La Defense Tel (1)42915613	
Termdok	Sur CD-Rom, accessible par PC, multilingue (anglais français, suédois, allemand, norvégien, danois, finnois et russe) 225000 termes avec definitions regroupe 7 banques de données terminologiques (Normaterm, Termium, TNC, Tepa,...)
Walters Lexicon C°, Södermalmstorg 8, 17800 Stockholm, Suède. tel +46(08)439510 US$ 920 ou AFNOR 6500FF HT	
Termnet	International Network for Terminology Production et diffusion de publications et de produits et services dans le secteur de la terminologie à l'échelle internationale
Heinestrasse 38, A-1020 Vienne, Autriche.	
Termium	Pour verifier et normaliser la terminologie dans les deux langues du Canada Pour aider les traducteurs dans leur travail
Université de Montréal ou Régie de la Langue Française du Québec.	
TDB	Terminology data bank Integree dans un système d'aide aux traducteurs et d'aide au développement de la terminologie
Carnegie Mellon University	

ANNEXE 2

Principaux logiciels d'OCR sur le marché français

(d'après 01 Informatique - N° 1084)

Logiciels	Editeurs/distributeurs	Prix
Autoread		
monoscanner	ISTC	6950 FF
multiscanner		8950 FF
Accatex	Datacopy/Alphasystem	9950 FF
Cognicar		
modèle 1	Cognisoft/Micropros	20000 FF
modèle 2		13900 FF
Discover 9320		
modèle 10	Kurzweil/Penta System	8000 FF
modèle 30		66000 FF
Image-Read		
image-in	CPI/MTE	4900 FF
K 5100		140000 FF
freedom	Kurzweil/Penta System	38000 FF
OCR +	Datacopy/Donatec	8950 FF
Omnipage 2.0	Caere/softmart	9150 FF
Readstar Express		9950 FF
Readstar D		4990 FF
Readstar 2 +	Inovatic	20000 FF
Readstar 3 +		40000 FF
Readstar 6		75000 FF
Readright 2.0	OCR System/Canon	4400 FF
Recognita		11900 FF
Recognita +	S2KI/Apsylog	10900 FF
Scaned	Calera/mentor Graphics	50000 FF
Texiris 2		49950 FF
Texiris 2 +	Iris/LCE	40000 FF
Text Pert 3.0	CTA/P Ingéniérie	9900 FF

Annexe 3: Rémunération des traducteurs

(d'après Loll Rolling, CCE, Luxembourg, 1982)

PAYS		FF/100 mots	
		Lang. Europ.	Lang. Exot.
Etats-Unis	A.T.A.	8,5	
	Free-lance	15 - 25	
Grande Bret.		15 - 30	
Belgique		20 - 28	36 - 80
Canada		21 - 64	
Suisse		33	
France		35 - 50	65 - 100
Suède		70	
Allemagne		65 - 100	

Eléments de comparaison (1987)

traduction brute...........14 FF/page de 250 mots
avec saisie................35....................
avec post-édition..........65....................
humaine..................160....................

Annexe 4: PRINCIPALES SOLUTIONS

Nom	Utilisation	Contact
Systèmes autonomes		
Systran 13%	USAF, Xerox, GM, WTC Canada, OTAN, CCE, Dornier, IONA, KFK, Aerospatiale, CEDOCAR	M. Loll Rolling M. I. Pigott ou Gachot S. A. 26 bis Av. de Paris B. P. 14 95230 Soisy s/s Montmorency - France
130000 lignes de programmes par langue, 100000 règles, 500000 mots/heure(théorie), 80000(pratique), IBM4381, 5 Gigaoctets, 20 centimes/mot(50FF/page)		
Logos 26%	CEE Luxembourg, Nixdorf, Opel, Siemens Mercedes, Pfaff, Burroughs	Logos Corp. 1, Dedham Place, Dedham, Ma. 02026 USA
Ang-vietnamien, français, allemand, espagnol IBM, Wang, Unisys...30 à 40 centimes/mot, y compris amortissement sur cinq ans.		
Metal	Siemens	PKI Philips Kommunications Industries
All-Angl, Angl-All.		
B'VITAL (Ariane)	SITE (France)	M.Pelletier CIGREF 21, Rue de Messine, 75008, Paris
1,5 million operations / mot en cours d'industrialisation, IBM 43XX, 30XX, 93XX		
Systèmes a syntaxe contrôlée		
TAUM		
TITUS	ITF Agriculture tropicale Allied Chemicals	M. J. M. Ducrot Institut Textile de France, 28, AV. A. Briand, BP 141 92223 Bagneux CEDEX France
Fran, angl, all, esp. Temps de reaction augmente de 10% pour écriture en langage TITUS. IBM Origine: prof. Baker USA.		
Systèmes interactifs		
Weidner 23%	Marine, Aérospatiale Bull...	TAO International, 37 ter, Rue de Metz 31000 Toulouse France
Transactive(ALPS) 12%	OTAN	
Ericsson 16%		
Grands projets:		
EUROTRA	13 Universités europ.	45 MECUS
CMT (E.U)	CMU (reconnaissance de la parole)	
ATR	Interp. téleph. Ang-Jap	4 Milliards FF
EDR	Electronic dictionary	1,5 ....

SOLUTIONS "JAPONAISES"(1)

(d'après JAPON IA JUIN 89 – lettre Ambassade de France/INIST)

Nom de la Compagnie	Nom du produit	Configuration de base	Matériel pour l'application (système d'exploitation)	Langages traduits	Date de sortie	Domaine	Types et tailles des dictionnaires (nbre de mots)	Possibilité de modif. du dict. par l'utilisateur	Vitesse de traduction (avec envirt matériel)	prix en yens	Ventes		Environnement
											Nbre d'utilisateurs	Nbre d'unités vendues	
Fujitsu Ltd.	Automatic Transl on System I (ATLAS I)	—	Moyen système à usage général	Anglais -> Japonais	?	Science et Technologie	- Dict. de base (53000) - Dict. Scientifique (250000)	oui	60000 mots/h Facom M-380	350000 par mois	150	—	Interactive system Batch system
	Automatic Transla ion System II (ATLA II)	Programme de traduction, Grammaire, Dictionnaire japonais, Dictionnaire anglais	Moyen système à usage général, série M (OSIV/F4 MSP)	Japonais -> Anglais	7/85	Science et Technologie	- Dict. de base (50000) - Dict. Scientifique (250000)	oui	60000 mots/h Facom M-380	550000 par mois	100	—	Interactive system Batch system
Toyo Information Systems Co., Ltd.	Denjirin (dictionnaire électronique)	Dictionnaires: Japonais-Anglais et Anglais-Japonais	Ordinateurs personnels serie PC9801, FM16 Beta, IBM5550 (MSDOS) 256 Ko nécessaires	Japonais ->Anglais, Anglais -> Japonais	12/85	Electronique, Electricité, Genie civil, Informatique, (13 domaines)	- Dict. de base (200000) - Dict. Scientifique (800000)	oui	—	98000	—	~ 1000	Ordinateur personnel
Toshiba Corporation	English to Japane e Translation Syste	—	Station de travail, UX-700 (UNIX) 4 Mo nécessaires	Anglais -> Japonais	?	Electronique, Mecanique, Chimie, Industrie, Informatique	- Dict. de base (30000) - Dict. utilisateur (50000) -Dict. spécialisé (50000)	oui	—	11450000	—	—	Interactive system Batch system
Datam System Co., Ltd	Hantran 2200	IBM5541-J08, Traitement de texte Coréen/Japonais	Ordinateurs personnels IBM5550 (MSDOS) 512 Ko nécessaires	Japonais -> Coréen	11/86	Manuels, articles...	- Dict. de base (66000)	oui	1 phrase/s	2400000 par mois	5	5	Batch system
Hitachi Ltd.	HICATS/JE	Programme de traduction, Dictionnaires, supports de post et pré-édition.	Moyen système à usage général, HITAC-M (VOS3) 4 Mo nécessaires	Japonais -> Anglais	5/86	Documents scientifiques et techniques, manuels...	- Dict. de base (50000) -Dict. spécialisé (250000)	oui	20000 à 60000 mots/h	550000 par mois	—	—	Interactive system Batch system
Sharp Corp.	English-Japanese Machine Translati n System	Programme de traduction, Dictionnaires	Station de travail, ix-5, ix-7, OA-90DX (UNIX System 5) 3 Mo nécessaires	Anglais -> Japonais	?	Documents techniques...	- Dict. de base (60000) - Dict. utilisateur (20000) -Dict. spécialisé (20000)	—	5000 mots/h (i-x-5, OA90-DX)	—	—	—	Interactive system Batch system

SOLUTIONS "JAPONAISES"(2)

(d'après JAPON IA JUIN 89 - lettre Ambassade de France/INIST)

Nom de la Compagnie	Nom du produit	Configuration de base	Matériel pour l'application (système d'exploitation)	Langages traduits	Date de sortie	Domaine	Types et tailles des dictionnaires (nbre de mots)	Possibilité de modif. du dict. par l'utilisateur	Vitesse de traduction (avec envirt matériel)	prix en yens	Ventes		Environnement
											Nbre d'utilisateurs	Nbre d'unités vendues	
Osaka Gas Information System	PENSEE	—	Ordinateurs personnels if1000 Unitopia 10M (Uniplus + System V) 8 Mo nécessaires	Japonais -> Anglais	1/87	Science et Technologie, Manuels	- Dict. de base (50000) - Dict. utilisateur (35000)	oui (dict. util.)	2000 mots/h if1000 Unitopia 10M	5600000	—	—	Interactive system
Oki Electric System	PENSÉE	—	Station de travail, if1000 (IF-UX) 88 Mo nécessaires	Japonais -> Anglais	2/87	Documents généraux, scientifiques et techniques	- Dict. de base (50000) - Dict. utilisateur (60000)	oui (dict. util)	4000 mots/h	6700000	—	—	Interactive system Batch system
NEC Corp.	PIVOT (Integrated Automatic Translation System)	Programme de traduction, Dictionnaire de base, Editeur pour le dict.	Moyen système à usage général, série ACOS (ACOS-4-MVP, ACOS-6-MVX, XE)	Japonais ->Anglais, Anglais -> Japonais	5/85	Science et Technologie	- Dict. de base (50000)	oui	132000 mots/h	- 490000 par mois ang->jap - 500000 jap->ang	—	5	Interactive system Batch system
Resource Sharing Company	STAR	Programme de traduction, Dictionnaire, Données grammaticales	Station de travail, Ustation E/20 EWS 4800 (UNIX) 21 Mo nécessaires	Anglais -> Japonais	2/87	—	- Dict. de base (10000) -Dict. spécialisé (15000)	oui	15000 mots/h 68020 cpu	11450000	—	—	Interactive system Batch system
Sanyo Electric Corp.	Word processor with Japanese-English translation facility	—	Trans-wordprocessor SWP-7800 (iRMX) 1 Mo nécessaire	Japonais -> Anglais	2/87	Général	- Dict. de base (50000) - Dict. utilisateur (100000)	oui	3500 mots/h	2380000	—	—	Appelé par le traitement de texte
SPIRIT, Inc	X-EJ	Programme de traduction, Dictionnaire de base, Editeur pour le dict.	Ordinateurs personnels serie PC-9800 (MSDOS) 256 Ko nécessaires	Anglais -> Japonais	4/86	—	- Dict. de base (14000) -Dict. utilisateur (12000)	oui (dict. util.)	—	98000	126	126	Interactive system Batch system

COMPARAISON ENTRE LES TRADUCTIONS HUMAINES ET LES TRADUCTIONS AUTOMATIQUES (QUALITES, COUTS ET DELAIS)

par

Oleg Lavroff
Société Aérospatiale
Chef du Centre d'Information Documentaire Commun
B.P. No. 76, 92152 Suresnes Cedex
France

Résumé

L'exposé présente des textes bruts traduits à l'aide de la machine (Traduction Automatique) et les textes post-édités (en version affinée) et indique les temps passés par un traducteur utilisant le système SYSTRAN. Un tableau récapitulatif fournit les temps et les coûts de traduction effectuée en TA et met en évidence les gain de productivité obtenus par rapport à une traduction totalement humaine. Les résultats présentés ne concernent que des traductions pour lesquelles la responsabilité juridique des Sociétés n'est pas engagée.

1. Considérations générales

La comparaison entre des traductions effectuées par un traducteur humain et celles obtenues par la TA (Traduction Automatique) soulève toujours des polémiques passionnées entre les traducteurs "classiques" (refusant la machine en tant que traducteur) etles traducteurs "nouvelle génération" (tirant un profit maximal de la machine).

Si l'on demande à un groupe, composé uniquement de traducteurs, de choisir la meilleure des traductions d'un texte effectuées par plusieurs traducteurs compétents dont l'une traduite par la machine et suivie d'un traitement de post-édition affinée réalisé par un adepte convaincu de la TA, l'expérience montre que ce groupe se trouve dans l'incapacité de sélectionner ce texte, voir même d'indiquer celui traité par la machine.

Au stade actuel de l'évolution des systèmes de TA, on peut affirmer, à condition de régler les problèmes de terminologie, d'analyser systématiquement les textes traduits, d'entretenir un dialogue permanent avec les concepteurs de systèmes et par ailleurs de faire appel à des traducteurs convaincus par la TA, que la qualité de traduction obtenue à l'aide de la machine et suivie d'une post-édition affinée est d'un niveau identique à celui obtenu par une traduction totalement humaine. Il faut cependant bien distinguer les domaines d'applications possibles et les nécessités éventuelles d'intégration des systèmes dans les Sociétés.

Dans la comparaison des textes traduits par un traducteur humain ou à l'aide de la machine on ne peut que comparer les résultats obtenus après une post-édition affinée qui, par définition, doit être équivalente à celle de la traduction humaine. Par conséquent, les exemples présentés au cours de cet exposé ne concernent que le texte source, la traduction texte machine et la post-édition affinée (la traduction humaine ne pouvant être que différente d'un traducteur à un autre).

Pour comparer en toute objectivité les deux modes de traduction, il faut impérativement faire appel dans le cas de la TA à des traducteurs motivés et objectifs et visant à obtenir une traduction de qualité humaine. Il faut par ailleurs, leur fournir tous les outils adaptés à leurs besoins (traitement de texte convivial, recherche terminologique intégrée au poste de travail, modem de connexion automatique au serveur du logiciel, etc ...)

Pour les textes nécessitant un très haut niveau de qualité, il est nécessaire pour la TA de faire intervenir d'autres facteurs techniques à mettre en jeu avant le lancement des traductions. Ainsi, par exemple, il faut faire appel à un correcteur orthographique (langue source), clarifier les ambiguïtés, réécrire si nécessaire les phrases trop longues et complexes, ressortir la terminologie inconnue dans le système. On arrive ainsi à définir en amont de la traduction des procédures de travail à respecter lors de la rédaction des textes. A ce jour, un bon nombre de spécialistes travaillent dans ce domaine en tenant compte du fait que de plus en plus les rédacteurs tentent de rédiger directement dans la langue cible. En conséquence, la comparaison des coûts/délais ne portera que sur des textes dits "d'information courante" devant être traduits et fournis rapidement.

Cette dénomination englobe d'une part, la notion "connaissance de l'information" pour laquelle on peut estimer qu'à 50 % des cas une traduction TA avec post-édition minimale est largement suffisante, et d'autre part, les textes diffusés à l'extérieur des Sociétés mais n'engageant pas en général leur responsabilité juridique.

Compte tenu de certains aspects techniques de réalisation de la documentation technique des Après-Vente il parait délicat à ce jour d'utiliser la TA dans ce domaine, à moins de disposer de logiciels pouvant facilement et économiquement s'intégrer dans les sites opérationnels des Sociétés.

2. Textes de comparaison

On trouvera en Annexe 1 trois textes de comparaison permettant d'illustrer le tableau des coûts et délais de traduction obtenue par la TA et effectuée par un traducteur humain.

Le premier texte est un extrait d'un compte rendu du Technical Committee on Technical Information de l'AIAA (traduction de l'anglais vers le français).

Le deuxième texte est un extrait d'une note technique traitant des techniques de contrôle non destructif (traduction du français vers l'anglais).

Le troisième texte est une note provisoire de travail relative à la préparation de notre cycle de conférences (traduction du français vers l'anglais).

Dans les trois cas la post-édition présentée est une post-édition affinée. Les textes ont été traduits à l'aide du système SYSTRAN à partir d'un poste de travail (micro type IBM PC) implanté dans une société.

Il est évident que le temps de post-édition affinée varie d'un texte à un autre et à l'intérieur même du texte, en fonction des domaines traités, de l'absence de terminologie déjà codée et en fonction de la rédaction des textes sources. En conséquence, à ce jour, le temps total de traduction indiqué sur ces exemples reflète un traitement minimal dans le meilleur des cas. Il peut se produire des cas où le temps de post édition affinée d'un paragraphe est supérieur au temps de traduction effectuée par un traducteur humain (en moyenne générale 250 à 300 mots à l'heure, selon les difficultés rencontrées).

3. Aspects économiques

Afin de raisonner en dehors de tout contexte monétaire les informations économiques sont fournies à partir des hypothèses et références suivantes pour une page de 250 mots :

- Traduction humaine
  - . Temps : 1 heure (frappe comprise)
  - . Coût : référence de base 100
- Traduction Automatique
  - Reconnaissance de caractères
    - Temps : 3,5 minutes
    - Coût OCR : 2,6 % par rapport à la référence de base.
  - Transmission, traitement et réception
    - Temps : 1,5 minute
    - Coût : 37,5 % par rapport à la référence de base.
  - Post-édition
    - . minimale
    - Temps : 10 minutes
    - Coût : 12,5 % par rapport à la référence de base.
    - . affinée
    - Temps : 33 minutes
    - Coût : 41,2 % par rapport à la référence de base.

- **Récapitulatif**

Traitement	Temps	Coût
Post-édition minimale	15 minutes	52,6% de la ref. de base
Post-édition affinée	38 minutes	81,3% de la ref. de base

Ces valeurs sont issues d'un bureau de traduction utilisant le système SYSTRAN à partir d'un poste de travail (micro type IBM PC) connectable à un serveur extérieur. Les statistiques sont basées sur environ 1000 pages portant sur des domaines techniques, économiques et de politique industrielle.

A ce jour, ce bureau traduit plus de 50 % des textes traduits en interne à l'aide de la T.A. et participe d'une manière très active à l'amélioration du système en transmettant au concepteur une analyse systématique des textes traduits.

Ces premiers résultats opérationnels très encourageants nous permettent d'établir le tableau récapitulatif suivant et de dresser un diagramme prévisionnel de l'évolution des coûts de la traduction automatique. Les valeurs du tableau sont données pour un lot de traitement de 10 pages (limite actuelle de transfert permettant de recevoir en ligne les traductions brutes machine).

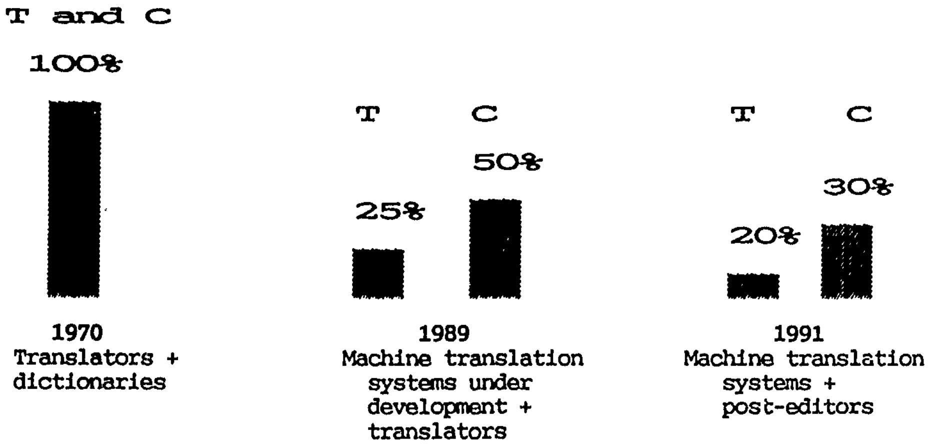
	OCR	Transfert et Traduc. machine	Post-édition		Total			
			minimale	affinée	P.-édit.mini.		P.-édit.affin.	
					10 p.	1p.	10 p.	1 p.
Temps	35 mn	15 mn	1h 40'	5h 30'	2h 30'	15mn	6h20'	38 mn
Coût(1)	2,6 %	37,5 %	12,5 %	41,2 %	-	52,6%		81,3 %

(1) par rapport à une référence de base traduction humaine :
. Temps : 250 mots par heure
. Coût : référence 100

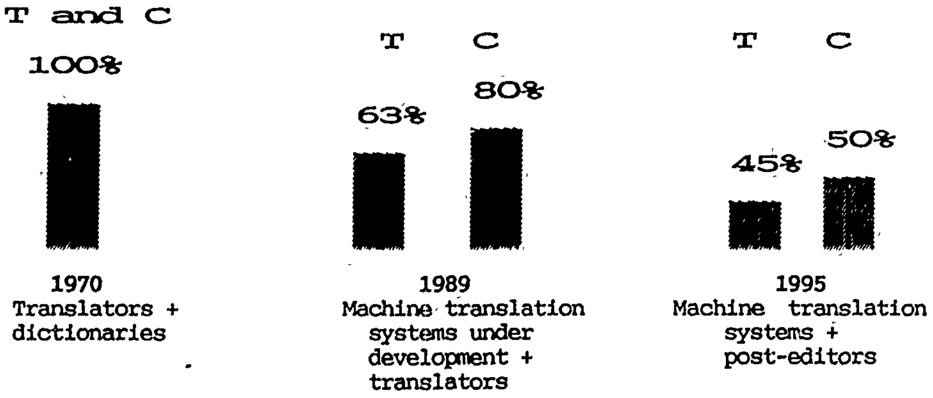
THE ECONOMIC FACTOR

A reduction in the costs and time of translation

1. Translation costs and time for informative texts (minimum post-editing)



2. Translation costs and time for documents to be dispatched abroad (refined post-editing)



L'investissement financier d'un poste de travail complet (voir les schémas ci-après) exprimé en francs français (FF) est le suivant :

. Un scanner et son logiciel de reconnaissance de caractères :

~ 80 000 FF

. Un micro type IBM PC comprenant un logiciel de traitement de texte convivial, une carte EGA, des cartes modem de liaison :

~ 50 000FF

. Une imprimante laser :

~ 20 000 FF

. Investissement total pour un poste de travail :

~ 150 000 FF

L'investissement pour plusieurs traducteurs est moins important du fait que le scanner et l'imprimante peuvent être partagés par les utilisateurs. Ainsi, pour un exemple de 5 traducteurs l'investissement par poste sera de l'ordre de 70 000FF. Dans ces conditions, l'amortissement financier pourra être réalisé en un an ou sur deux années environ pour un seul poste.

4. Conclusions

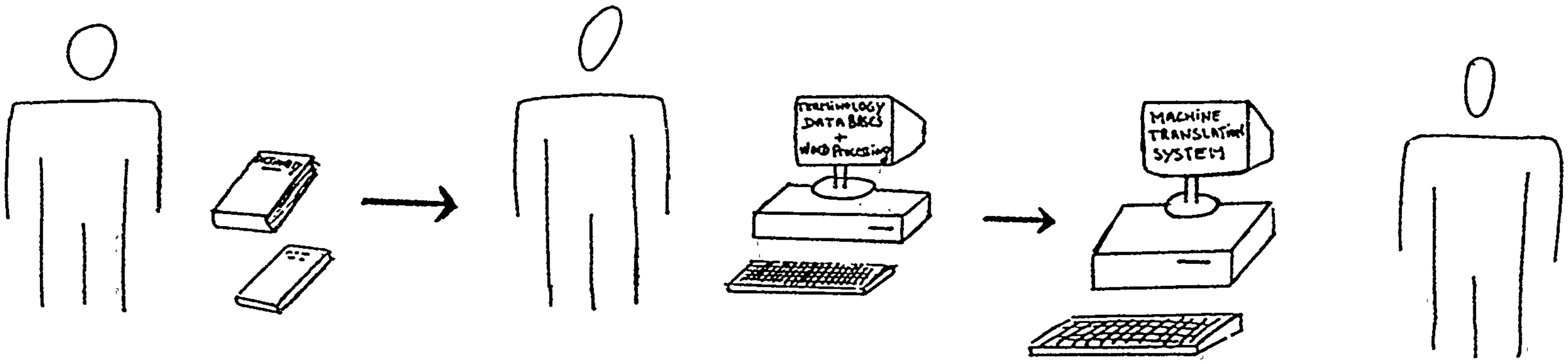
On peut estimer à ce jour que l'introduction de la Traduction Automatique dans les bureaux de traduction, à condition de satisfaire à toutes les exigences humaines et matérielles, et de promouvoir correctement cette nouvelle technique de traduction, devrait permettre à court terme d'améliorer notablement la productivité de ces bureaux.

L'avènement de la TA nous amène à redéfinir les tâches du traducteur et à transférer vers les secrétariats des travaux qui ne nécessitent pas la compétence des traducteurs (par exemple, la reconnaissance des caractères). Ainsi, les systèmes de TA permettent à ce jour, pour des domaines et des applications bien définies, de traiter environ deux pages à l'heure en post édition affinée. Certains spécialistes ou concepteurs de systèmes estiment qu'il est possible de traiter ainsi 3 pages à l'heure. Pour notre part, nous estimons que, dans l'état actuel des choses, 2 pages de post édition par heure nous semblent tout à fait réalisables, ce qui nous amène à conclure que, dans ces conditions, le gain potentiel de la TA est de l'ordre de :

- 37% sur les temps de traitement
- 20% sur les coûts.

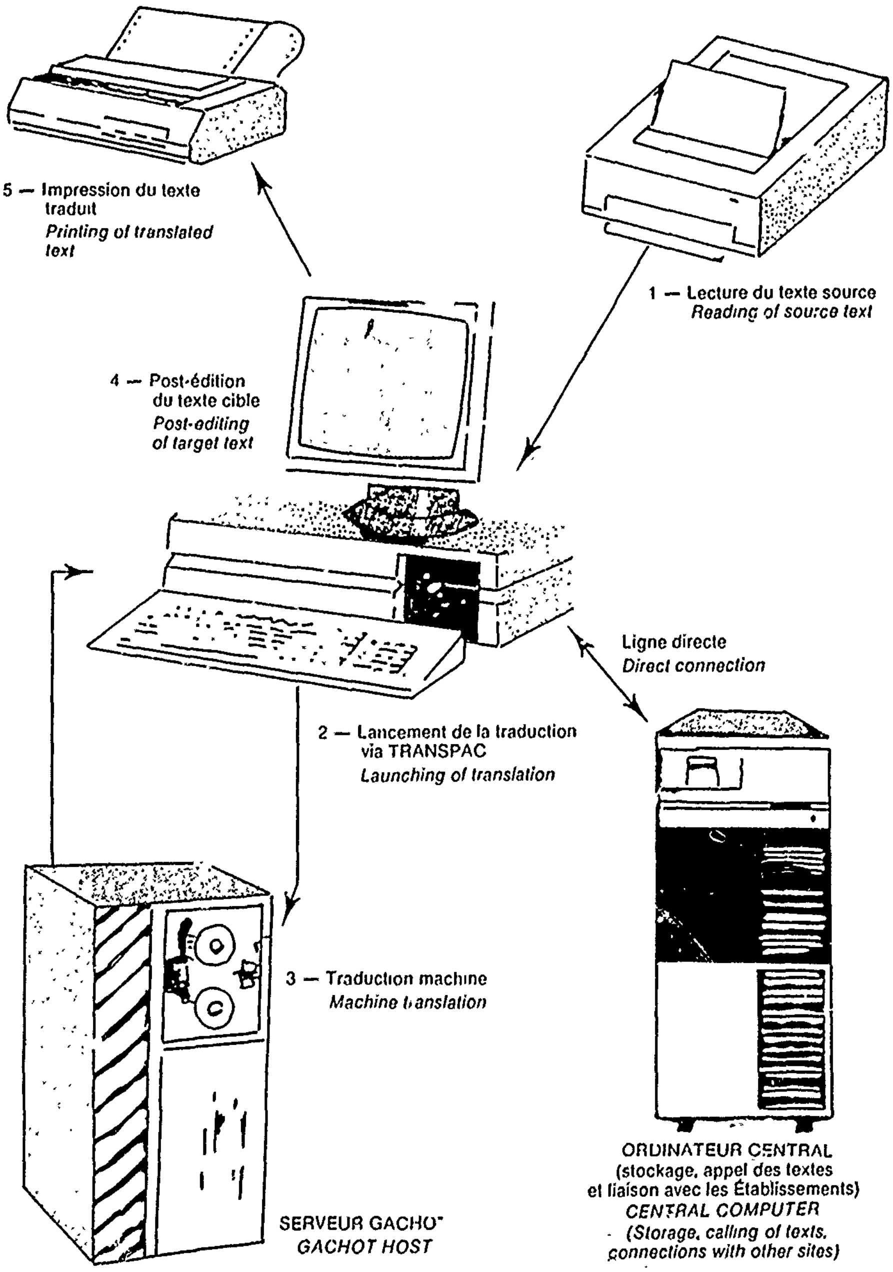
THE HUMAN FACTOR

The evolution of the job of translators


TERMINOLOGY DATA BASES + WORD PROCESSING
MACHINE TRANSLATION SYSTEM

Yesterday's translator	Today's translator	Tomorrow's translator: **The Post-Editor**
1970	1988	1990

MODE DE FONCTIONNEMENT
OPERATING CHART


5 — Impression du texte traduit
Printing of translated text
1 — Lecture du texte source
Reading of source text
4 — Post-édition du texte cible
Post-editing of target text
Ligne directe
Direct connection
2 — Lancement de la traduction via TRANSPAC
Launching of translation
3 — Traduction machine
Machine translation
SERVEUR GACHOT
GACHOT HOST
ORDINATEUR CENTRAL
(stockage, appel des textes
et liaison avec les Établissements)
CENTRAL COMPUTER
(Storage, calling of texts, connections with other sites)

Comparison between human and automatic translations (quality, costs and processing time)

MR. O. LAVROFF

-=-=-=-=-=

Abstract

In this paper, rough machine translations or automatically translated texts and post-edited texts (resulting from a refined post edition) will be presented together with the respective time a translator devotes to translation when using the SYSTRAN system. The times and costs of automatic translation are summarized in a chart which thus highlights the increased productivity of CAT compared to an all-human translation. The results mentioned hereafter only apply to translations for which the company's liability is not involved.

1. General overview.

The comparison between texts translated by a human translator and rough machine translations always raises an impassioned controversy between "classical" translators (who deny the use of a machine as translators) and the "new generation" translators (who take the most of the machine).

If a group only consisting of translators is asked to choose the best translation of a text among a number of translations performed by several qualified translators and one text translated by the machine and thoroughly post-edited by an advocate of automatic translation, the experience shows that such a group is unable to select the best translated text or even to find out the text resulting from an automatic translation system.

At the current development stage of automatic translation systems, it can be asserted that the quality of a rough machine translation thoroughly post-edited is similar to that of an entirely human translation, provided all the problems related to terminology have been settled, the rough machine translations are systematically analyzed, a permanent contact is kept with system manufacturers and furthermore, provided the translators involved are convinced of the benefits of automatic translation system. A clear distinction shall however be made between the possible fields of application and the possible requirements of integration of such systems in the companies.

When comparing texts translated by a human translator with texts translated with a machine, the only relevant terms of comparison are the results of a machine translation followed by a refined post-editing with a quality equivalent to that of a human translation. Consequently, examples given in this paper only concern the source text, the machine translated text and the refined post-editing (bearing in mind that a human translation is different from one translator to another).

In order to compare objectively the two translation types, motivated and objective translators have to be called upon as far as machine translation is concerned. They have to strive to achieve a translation with the quality of a human translation. Besides, translators have to be provided with tools adapted to their needs (user-friendly word processing, integrated terminology search system connected to a word processing, automatical modem cards for the connection with the host system, etc...).

For texts requiring a high quality level, other technical factors have to be examined before sending the text for machine translation. For example, a spelling corrector has to be used, ambiguities clarified, long or complex sentences rewritten, terminology unknown to the system identified. It is then possible to define, upstream from translation, working procedures while writing down texts. To date, several specialists are working in this field, taking into account that more and more redactors try to write directly into the target language. Consequently, the comparison between cost and time will only concern the so-called "common information" texts, which require rapid translation and supply.

This heading means, on the one hand, the "information knowledge" idea, where it can be estimated that for around 50% of texts, a machine translation with a minimum post-editing is largely sufficient, and on the other hand, texts disseminated outside companies but for which their liability is not involved.

If we consider some technical aspects of after-sales technical documents, it seems to date tricky to use machine translation in this field, unless software can, easily and without undue expenses, be integrated in the companies operational sites.

2. Comparison of texts.

Appendix 1 shows 3 texts for comparison, which permit to illustrate the cost and time chart for machine translation and for human translation.

The first text is an extract from a Technical Information Technical Committee Report of the AIAA (translation from English into French).

The second text is an extract from a technical memorandum on non destructive techniques (translation from French into English).

The third text is a background paper on the preparation of our Lecture Series (translation from French into English).

In all cases, the post-edited version shown is a refined post editing. Texts were translated by the SYSTRAN system from a work station (such as an IBM-PC) used within a company.

It is clear that the time required for a refined post-editing varies from one text to another and inside the text itself, depending on the subjects treated, the terminology already coded and the quality in writing of the source text. Consequently, to date, the total translation time indicated represents the minimum processing for the best possible result. It may happen sometimes that the post-editing time spent for a paragraph is superior to that of a human tanslation (that is an average of 250/ 300 words per hour, depending on the difficulties encountered).

3. Economic aspects

In order to leave aside any currency aspect, the economic information are given from the following hypotheses and references for a 250 word page :

-Human translation

- . Time : 1 hour (typing included)
- . Cost : basic index 100

- Machine translation

. Optical character reading

- Time : 3.5 min.
- Cost : 2.6% of basic index

. Transmission, processing and reception :

- Time : 1,5 min.
- Cost : 37.5 % of basic index

. Post-editing

. minimum

- Time : 10 min
- Cost : 12.5 % of basic index

. refined
- Time : 33 min.
- Cost : 41.2 % of basic index

. Summary

Processing	Time	Cost
Minimum post-editing	15 Min	52.6 % of basic index
Refined post-editing	38 Min.	81.3 % of basic index

Theses values have been delivered by a translation bureau using the SYSTRAN system at a work station (such as an IBM-PC), connectable to an external host sytem. Statistics have been made from about 1000 pages, concerning technical, economic and industry subjects.

To date, this bureau translates over 50% of in-house translated texts by machine and participates very actively to the improvement of the system by sending the designer a systematic analysis of translated texts.

These first operational results, very promising, enable us to draw the following summary chart and the prospective diagram of machine translation development. The chart values are given for a batch of 10 pages (current tranfer limit without hindering the on-line reception of the rough machine translation).

	OCR	Transfer and Machine translation	post editing		total			
			minimum	refined	minimum		refined	
					10p.	1p.	10p.	1p.
Time	35'	15'	1hr 40'	5hrs 30'	2hrs30'	15'	6hrs20'	38'
Cost (1)	2.6%	37.5%	12.5%	41.2%	-	52.6%	-	81.3%

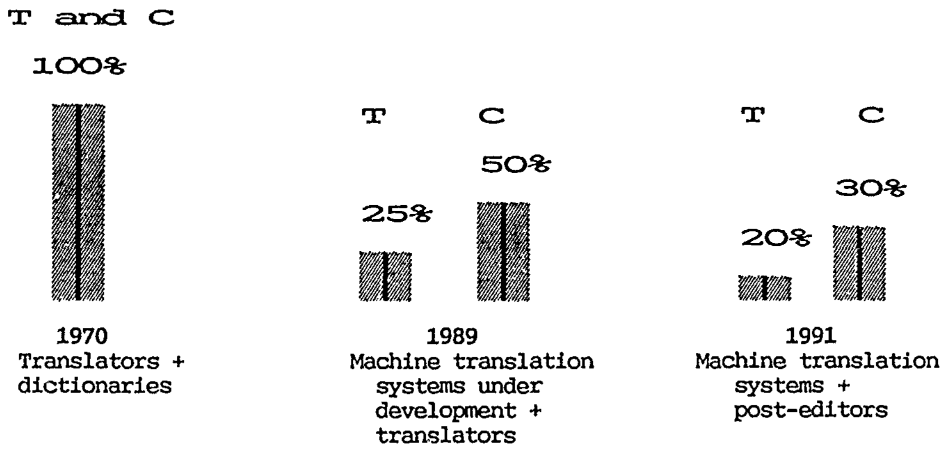
(1) with reference to an average basic human translation :
. time 250 words per hour
. Cost : index 100

THE ECONOMIC FACTOR

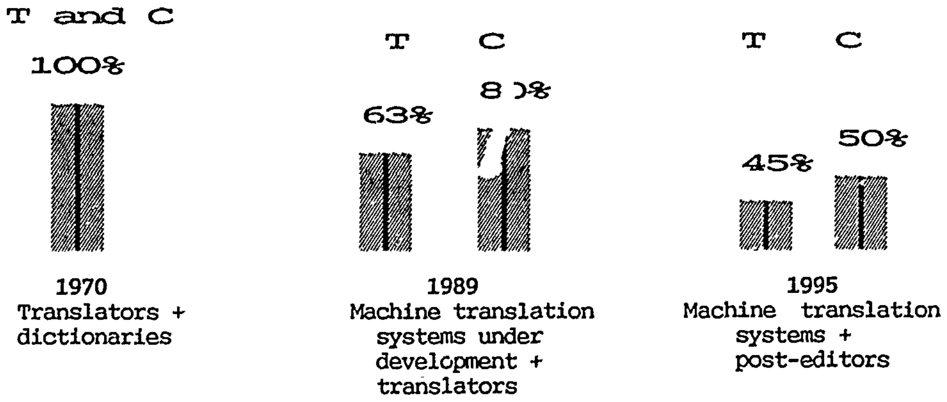
A réduction in the costs and time of translation

---

1. Translation costs and time for informat texts (minimum post-editing)



2. Translation costs and time for document to be dispatched abroad (refined post-editing)



The financial investment for a full work station (see following diagrams) expressed in French Francs, is the following :

. a scanner and an optical reading software :

80,000 FF

. a micro computer, such as an IBM-PC, equipped with a user-friendly word processing software, an EGA card and modem cards for connection :

50,000 FF

. a laser printer

20,000 FF

. Total investment for one work station

150,000 FF

The investment for several translators is reduced, as the scanner and the printer can be shared between users. For example, for 5 translators, the investment per work station will amount to about 70,000 FF. In these conditions, the investment could be amortized within one year or within two years or so for a single work station.

4. Conclusion

To date, the introduction of machine translation in translation bureaux should, in the short run, provided all human and material requirements are met and this new translation technique is correctly promoted, improve significantly the productivity of these bureaux.

The emergence of machine translation leads to a redefinition of the translator's tasks and to a transfer to secretaries of tasks which do not require translator skills (for example, optical character reading). Thus, machine translation systems permit, to date, for well defined fields and applications, to process around 2 refined post-edited pages per hour. Some specialists or system designers believe that 3 pages may be processed per hour. We think, however, that for the time being, 2 post-edited pages per hour are perfectly achievable, which brings us to conclude that, in these cicumstances, the potential gain of machine translation is about :

- 37 % of processing time
- 20% of cost.

Appendix 1

TEXT N°1

"AIAA Technical Committe on Technical Information
Minutes of June 13, 1989, Meeting"

I LANGUAGE PAIR :

English into French

II GLOSSARIES SELECTED :

- Aviation and Space
- Legal
- Political Science

III SUCCESSIVE OPERATIONS OF THE MACHINE TRANSLATION PROCESS :

- Optical reader --> Time : 5 mn
- Sending a text for translation --> Time : 4 mn
- Post editing / Minimum --> Time : 5 mn
  / Refined --> Time : 15 mn

- Number of words translated --> 361

IV ANALYSIS OF THE ROUGH MACHINE TRANSLATION :

a/ Terminology

Source text	Rough machine translation	Human translation
to issue minutes	établir des minutes	publier le compte rendu
current (members)	(membres) courants	(membres) actuels
information flow	écoulement de l'information	circulation de l'information
a topic	une matière	un sujet
information related associations	associations reliées par information	associations liées au secteur de l'information
companies	compagnies	sociétés
programs	régimes	programmes
(articles) highlighting sth	accentuant qch	relatifs à qch
to be due for (publication)	être dû pour la publication	seront publiés

a/ Terminology (ctd)

Source text	Rough machine translation	Human translation
charter	charte	statuts
a discussion was held	une discussion a été maintenue	une discussion a eu lieu
survey	aperçu	enquête
there was agreement to	il y avait convention pour	il a été convenu de
display (areas)	visualisation	exposition

b/ Grammar

Source text	Rough machine translation	Human translation
sb will be contacted on	qqun sera entré en contact pour	qqun sera chargé de
it was agreed that	il a convenu que	il a été convenu que
(that the TC) appoint	(que le TC) nomment	(que le TC) nommera
by encouraging critical evaluations	par des évaluations critiques d'encourager	en encourageant des évaluations critiques

c/ Defective analysis / prepositions

Source text	Rough machine translation	Human translation
Meeting	Réunissant	Réunion
Since X is the (beginning)	Depuis X a lieu le (commencement)	Etant donné que X est
(on the use and) mailing	(sur l'utilisation et) expédier	d'utiliser et d'expédier
(liaison) to sb	(liaison) à qqun	(liaison) avec qqun
on writing sth	sur écrire qch	au sujet de l'élaboration de qch
input (from)	entré	les résultats
if appropriate	si approprié	s'il y a lieu

e/ Word order

Source text	Rough machine translation	Human translation
Technical Committee	Technique Comité	Comité Technique
Electronic Transfer of Information	Transfert de l'information électronique	Transfert Electronique de l'Information
(are due) September 7	(sont) le 7 septembre dû	seront publiés le 7 septembre
scope statement	le rapport de portée	domaine d'application
additional members and ... vendors	membres et fournisseurs additionnels	membres supplémentaires et fournisseurs

f/ Unrecognized words

Source text	Rough machine translation	Human translation
AIAA	%AIAA	l'AIAA
TCs	%TCs	Comités Techniques
TCTI	%TCTI	CTIT

SOURCE TEXT N°1

AIAA Technical Committee on Technical Information

Minutes of June 13, 1989, Meeting

New York

Since May/June is the beginning of the TC cycle, B. Lawrence said that this second meeting of the TC was appropriate to issue formal minutes with copies to selected AIAA headquarters staff (1). A list of current TC members will be submitted to headquarters (2).

R. Lewis suggested that workshops be sponsored or special studies be undertaken on the processes of information flow and specific technologies such as CD-ROM or electronic publishing. H. Mindlin said that ASME has a database committee which promotes this type of activity and represents different disciplines; a similar group may work within the AIAA structure. The topic selected is "Electronic Transfer of information and Its Impact on Aerospace Research and Development". Membership lists from information related associations will be used to target managers in local aerospace companies and organizations (8). AIAA headquarters will be contacted on the use of a conference room and mailing the brochures (9).

TC members discussed the importance of collaborating with other TCs on programs and activities. It was agreed that the TC appoint a person as liaison to the Publications Committee.

Article for Aerospace America. Articles from each TC highlighting the year's activities by discipline are due September 7 for publication in Aerospace America. The content of our TC's submission will be on developments in aerospace technical information (14).

TC Charter. A brief discussion was held on writing a charter for the TC. Since input from the survey will help determine the role of the TC, there was agreement to expand the scope statement provided by B. Lawrence in her letter of March 23, 1988.

The Technical Committee on Technical Information (TCTI) promotes the development of aerospace scientific and technical information services. The TCTI encourages the flow of technical information throughout the aerospace community by organizing activities which provide a forum for the exchange of ideas and by encouraging critical evaluations of information transfer processes.

Closing Remarks. Items on the agenda not discussed were recommendations of additional committee members and information vendors in display areas. If appropriate, both will be discussed at the next TC meeting.

ROUGH MACHINE TRANSLATION N°1

```
MOTS TRAITES :     361 COMPTE AVANT :     286163- COMPTE APRES :
    286524-
0100100000PDEM    EF     YPA TG=4LPDEBUG=S SYS=UCDATE=19 02 90
11H15    0002544
```

Technique comité %AIAA d'information technique

Compte rendu du 13 juin, 1989, réunissant

New York

Depuis mai/juin a lieu le commencement du cycle de TC, B. Laurent dit que cette deuxième réunion du TC était appropriée pour établir des minutes formelles avec des copies au personnel sélectionné des sièges sociaux %AIAA (1). Une liste des membres courants de TC sera soumise aux sièges sociaux (2).

R. Lewis a proposé que des ateliers soient pris en charge ou des études spéciales soient entreprises sur les procédés de l'écoulement de l'information et des technologies spécifiques telles que la CD-$CROM ou l'édition électronique. H. Mindlin a dit qu'ASME a un comité de base de données qui favorise ce type d'activité et représente des disciplines différentes; un groupe semblable peut travailler dans la structure %AIAA. La matière sélectionnée est " transfert de l'information électronique et son impact sur la recherche et le développement aérospatiaux ". Des listes des membres des associations reliées par information seront employées pour viser des directeurs dans les compagnies et les organismes aérospatiaux locaux (8). Les sièges sociaux %AIAA seront entrés en contact sur l'utilisation d'une salle de conférence et expédier les brochures (9).

Les membres de TC ont discuté l'importance de la collaboration avec l'autre %TCs sur des régimes et des activités. Il a convenu que le TC nomment une personne comme liaison au Comité de publications.

Article pour l'Amérique aérospatiale. Les articles de chaque TC accentuant les activités de l'année par discipline sont le 7 septembre dû pour la publication en Amérique aérospatiale. La teneur de la présentation de notre %TC sera sur des développements dans l'information technique aérospatiale (%14) .

Charte de TC. Une brève discussion a été maintenue sur écrire une charte pour le TC. Puisqu'entré de l'aperçu aidera à déterminer le rôle du TC, il y avait convention pour augmenter le rapport de portée fourni par B. Laurent dans sa lettre du 23 mars, 1988.

Le comité technique de l'information technique (%TCTI) favorise le développement des services d'information scientifique et technique aérospatiale. Le %TCTI encourage l'écoulement d'information technique dans toute la communauté aérospatiale par l'organisation des activités qui fournissent un forum pour l'échange des idées et par des évaluations critiques d'encourager des procédés de transfert de l'information.

Observations finales. Les articles aux ordres du jour non discutés étaient des recommandations des membres de comité et des fournisseurs additionnels de l'information dans des zones de visualisation. Si approprié, tous les deux seront discutés lors de la prochaine réunion de TC.

POST-EDITING N°1

Comité Technique de l'AIAA sur l'Information Technique

Compte Rendu de la Réunion du 13 juin 1989

New-York

Etant donné que les mois de Mai et de Juin correspondent au début du cycle de réunions du Comité Technique, Mme B. Lawrence a estimé que le moment était opportun de publier le compte rendu officiel et d'en envoyer un exemplaire à certains membres des sièges sociaux de l'AIAA (1). Une liste des membres actuels du Comité Technique sera soumise au siège social (2).

M. R. Lewis a proposé d'organiser des ateliers ou de procéder à des études spéciales sur les moyens de circulation de l'information et les techniques spécifiques telles que le CD-CROM ou l'édition électronique. M. H. Mindlin a déclaré qu'au sein de l'ASME, une commission spécialisée dans les bases de données encourage ce genre d'activité et agit dans diverses disciplines. Un groupe semblable pourrait travailler au sein de l'AIAA.
Le sujet retenu est "Le Transfert Electronique de l'Information et ses Répercussions sur la Recherche et le Développement dans le Domaine Aéronautique et Spatial".
Des listes des membres des associations liées au secteur de l'information seront utilisées pour localiser les directeurs des sociétés et organismes régionaux dans le secteur aéronautique et spatial (8). Le siège social de l'AIAA sera chargé de trouver une salle de conférence et d'expédier les brochures (9).

Les membres du Comité Technique ont discuté de l'importance de la collaboration avec d'autres Comités Techniques pour les programmes et les activités. Le Comité Technique nommera une personne pour assurer la liaison avec le Comité de Publications.

Articles pour Aerospace America. Les articles de chaque Comité Technique relatifs aux activités de l'année par discipline seront publiés le 7 septembre. Les conclusions de notre Comité Technique porteront sur le développement de l'information technique dans le monde aéronautique et spatial.

Statuts du Comité Technique. Une brève discussion a eu lieu au sujet de l'élaboration de statuts du Comité Technique. Etant donné que les résultats de l'enquête aideront à déterminer le rôle du Comité Technique, il a été convenu d'élargir le domaine d'application défini par Mme .B. Lawrence dans sa lettre du 23 mars 1988.

Le comité technique de l'information technique (CTIT) encourage le développement des services d'information scientifique et technique dans le domaine aéronautique et spatial. Le CTIT promeut également la circulation de l'information technique dans toute la communauté aérospatiale en organisant des activités qui favorisent les échanges d'idées et en encourageant les évaluations critiques des procédés de transfert de l'information.

Observations finales. Les deux articles inscrits à l'ordre du jour et qui n'ont pas été discutés concernaient des recommandations des membres supplémentaires de la commission, et des fournisseurs d'information lors d'expositions. S'il y a lieu, ces deux articles feront l'objet d'une discussion lors de la prochaine réunion du Comité Technique.

TEXT N° 2

"Non destructive testing techniques"

I LANGUAGE PAIR :

French into English

II GLOSSARIES SELECTED :

- Aviation and Space
- Chemistry
- Mechanical engineering

III SUCCESSIVE OPERATIONS OF THE MACHINE TRANSLATION PROCESS :

- Optical reader --> Time : 6 mn
- Sending a text for translation --> Time : 6 mn
- Post editing / Minimum --> Time : 10 mn
  / Refined --> Time : 20 mn
- Number of words translated --> 838

IV ANALYSIS OF THE ROUGH MACHINE TRANSLATION :

a/ Terminology

Source text	Rough machine translation	Human translation
Assemblages par collage	Assemblies by joining	Bonded assemblies
(Faire un) pas	(To take a) pitch	(To make a) step forward
Cuisson	Cooking	Curing
Atout	Trump	Asset
Système de recopie	Output system	Printout system
Se tourner vers	To turn to	To turn towards
Mettre en place	To install	To set up
(Produit) réalisé	(Product) carried out	(Product) manufactured
Procéder à des mesures	To carry out measurements	To perform measurements
Décollements	Separations	Debondings
Liquide de couplage	Fluid of coupling	Couplant fluid

b/ Defective analysis / prepositions

Source text	Rough machine translation	Human translation
Un (éventail)	One (range)	A (range)
Des (problèmes)	Of the (problems)	__ (problems)
Des meilleures techniques	Best techniques	The best techniques
De nombreuses	The many	Many
Que ce soit...	Whether it is	Either ... or
Aux (USA)	To (the USA)	In (the USA)
Expérience de	Experience of	Experience in
Un ensemble de	A whole of	Many
Les besoins de (quelqu'un)	The needs for (sb)	The needs of (sb)

c/ Word order

Source text	Rough machine translation	Human translation
Les industries automobile, aéronautique et même électronique	Car industries, aeronautical and even electronic	The aerospace, car and even electronics indutries
Les méthodes les plus adaptées	The methods the most adapted	The most adapted methods
Des différents partenaires	Different the partners	The different partners
Concurrence étrangère	Competition foreign	Foreign competition
Liés justement à	Connected precisely with	Precisely connected with
Ondes de cisaillement ou de lamb	Waves of shearing or lamb	Shear or lamb waves
Analyse modale	Analyzes modal	Modal analysis
Détectés rapidement	Detected quickly	Quickly detected

d/ Unrecognized words

Source text	Rough machine translation	Human translation
Multipartenaire Partenaire	%multipartenaire Partenaire	Multipartner Partner

V FINAL REMARKS

As far as French-speaking companies such as Aérospatiale are concerned, the documents to be translated into English or any other non-French language are generally designed to be dispatched abroad and thus require a refined post-editing of the rough machine translation. In such cases, the intervention of a human translator is necessary but undoubtedly remains quicker and cheaper than in an entirely human translation process.

SOURCE TEXT N°2

Les techniques de contrôle non destructif

Un très large éventail d'industries utilise les assemblages par collage mais elles sont limitées dans leurs applications du fait d'un manque de techniques de contrôle non destructif permettant de détecter des défauts pouvant limiter la fiabilité et la durée de vie de la structure. Ces défauts sont avant tout des problèmes d'adhésion entre la colle et la pièce et des problèmes de cohésion (qualité de la colle) qu'il faut détecter et évaluer d'une manière non destructive. Le but de ce projet est de rassembler des compétences européennes afin de faire un pas décisif dans le contrôle de ces structures (et des interfaces en général) en mettant au point de nouvelles techniques ultrasonores permettant d'améliorer la rapidité et les capacités de détection. Ces recherches forment un programme complet des meilleures techniques que l'on peut envisager dans ce domaine.
L'interconnexion entre ces différents laboratoires permettra de comparer et de compléter les techniques qui seront alors évaluées sur un .round .robin .test. Les répercussions sont importantes dans les industries automobile, aéronautique et même électronique.

Si de nombreuses industries sont tentées d'utiliser les assemblages par collage, elles se heurtent le plus souvent au problème de l'assurance de la qualité du produit final. Cette qualité se base sur la maîtrise des procédés de fabrication mais aussi sur un contrôle non destructif capable de mettre en évidence les défauts pouvant limiter la durée de vie de la structure. Ces défauts de type collectif (porosité, mauvaise cuisson) ou de type adhésif (absence de contact ou contact sans adhésion) peuvent se produire en cours de fabrication et se dégrader en cours de service. Les méthodes les plus adaptées pour la détection de ces défauts sont avant tout des techniques ultrasonores. Tous les partenaires de ce projet ont déjà une sérieuse expérience dans ces techniques, que ce soit du point de vue de la recherche, du développement ou de l'utilisation d'appareils déjà commercialisés par les partenaires 1 et 3.
Le regroupement des compétences européennes permettra de faire une évaluation comparative de nouvelles techniques novatrices en se basant sur les réflexions des différents partenaires ainsi que des études menées actuellement aux Etats-Unis.

Tous les partenaires de ce projet ont déjà une sérieuse expérience technique dans le sujet et nous pouvons assurer que ce programme permettra d'obtenir des résultats tout à fait satisfaisants. De plus, les Partenaires 1, 2 et 3 ont également une grande expérience de programmes multipartenaires ce qui est un atout supplémentaire pour le succès de ce projet.

- Réduction des temps de contrôle

- Faire face à la concurrence étrangère en proposant des produits plus fiables et mieux conçus, que ce soit dans le domaine aéronautique, automobile ou électronique.

- Augmentation de la fiabilité par l'emploi d'un système de recopie (automatique ou semiautomatique) connecté avec un système expert limitant l'interprétation humaine.

- Meilleure connaissance des comportements des colles et des joints collés conduisant à une utilisation plus rationnelle de ce procédé d'assemblage.

- Gain de poids du fait de la possibilité d'utiliser des renforts locaux là où les contraintes sont importantes.

6.1 Présentation générale

L'industrie moderne se tourne de plus en plus vers l'utilisation de matériaux composés de couches ou de protections successives et qui nécessitent de mettre en place des méthodes non destructives de contrôle afin d'assurer la qualité des interfaces et du produit fini. Ceci est particulièrement vrai pour les assemblages par collage qui présentent de nombreux avantages (réduction de poids, meilleure répartition des contraintes) et qui permettent des conceptions de structures ou des positionnements impossibles à réaliser par soudage ou rivetage.
Cependant l'utilisation intensive de ces assemblages se heurte à des problèmes liés justement à l'assurance de la qualité du produit réalisé.

6.2 Contenu scientifique du projet

Pour atteindre les objectifs décrits dans le paragraphe 2 et pour résoudre les problèmes exposés ci-dessus, nous mettrons au point un ensemble de nouvelles techniques ultrasonores.
Des aspects fondamentaux seront abordés en étudiant les propriétés des câbles et des assemblages dans le but de répondre aux besoins des bureaux d'études pour le dimensionnement et la compréhension des phénomènes de dégradation à l'échelle micrométrique (Partenaires 1 et 4).
Des études seront menées pour la détection des défauts d'adhésion en utilisant des ondes de cisaillement ou de .lamb sur une large gamme de fréquences (de 1 à 100 Hz), (Partenaires 3 et 5).

Des essais en vibration seront effectués soit pour procéder à des mesures plus globales sur une structure (analyse modale, Partenaire 2) ou encore pour créer des décollements sur les zones de mauvaises adhésions pour un .proof .test ultrasonore (Partenaire 6). Ces défauts pourront alors être détectés rapidement avec des capteurs rotatifs sans liquide de couplage (Partenaire 1 avec support théorique du Partenaire 7).

Toutes ces études seront menées dans le but de constituer un système expert à partir de l'extraction des informations représentatives des défauts recherchés (Partenaire 2 en collaboration avec les autres Partenaires).

ROUGH MACHINE TRANSLATION N° 2

```
MOTS TRAITES :     838 COMPTE AVANT :      252161- COMPTE APRES :
    252999-
0100100000P5123  FE      YPA TG=495DEBUG=S SYS=UCDATE=18 10 89
14H32    0005645
```

Techniques of non destructive testing

One very broad range of industries uses the assemblies by joining but they are limited in their applications because of a lack of techniques of non destructive testing making it possible to detect defects being able to limit the reliability and the lifespan of the structure. These defects are above all problems of adhesion between the adhesive and the part and of the problems of cohesion (quality of the adhesive) which it is necessary to detect and evaluate in a non destructive way. The aim of this project is to gather European competences in order to take a decisive pitch in the control of these structures (and of the interfaces in general) by developing the new ultrasonic techniques making it possible to improve the speed and the capacities of detection. This research forms a complete program of best techniques than one can consider in this field. The interconnection between these various laboratories will make it possible to compare and supplement techniques which will then be evaluated on a round robin test. The repercussions are significant in the car industries, aeronautical and even electronic.

If the many industries are tempted to use the assemblies by joining, they generally run up against the problem of the quality assurance of the end product. This quality is based on the control of the methods of manufacture but also on a non destructive testing able to highlight the defects being able to limit the lifespan of the structure. These defects of the collective type (porosity, the bad cooking) or of adhesive type (absence of contact or contact without adhesion) can occur in the course of manufacture and be degraded in the course of service. The methods the most adapted for detection of these defects are above all ultrasonic techniques. All the partners of this project have already serious experience in these techniques, whether it is from the point of view of research, development or use of aircraft already marketed by partners 1 and 3.
The regrouping of European competences will make it possible to make a comparative evaluation of the new innovative techniques while being based on the reflexions of different the partners as well as studies currently undertaken to the United States.

All the partners of this project have already serious technical experience in the subject and we can ensure that this program will make it possible to obtain completely satisfactory results. Moreover, Partners 1, 2 and 3 also have great experience of programs %multipartenaires what is an additional trump for the success of this project.

- Reduction of times of control

- Deal with the competition foreign by proposing more reliable and better designed products, either in the aeronautical field, automobile or electronic.

- Increase in reliability by the use of an output system (automatics or semiautomatic) connected with an expert system limiting human interpretation.

- Better knowledge of the behaviors of the adhesives and the adhesive bonded joints leading to a more rational use of this method of assembly.

- Gain of weight because of the possibility of using local reinforcements where the stresses are significant.

6,1 General Presentation

Modern industry turns more and more to the use of materials made up of layers or successive protections and which require to install non destructive methods of control in order to ensure the quality of the interfaces and finished product. This is particularly true for the assemblies by joining which have many advantages (reduction of weight, better distribution of the stresses) and which allow designs of structures or positionings impossible to realize by welding or riveting.
However the intensive use of these assemblies encounters problems connected precisely with the quality assurance of the product carried out.

6,2 Scientific Contents of the project

To achieve the goals described in paragraph 2 and to solve the problems mentioned above, we will develop a whole of the new ultrasonic techniques.
Fundamental aspects will be approached by studying the properties of the cables and the assemblies with the aim of meeting the needs for the design offices for the dimensioning and the comprehension of the phenomena of degradation on a micrometric scale (Partners 1 and 4) .
Studies will be undertaken for the detection of the defects of adhesion by using waves of shearing or lamb on a broad frequency range (from 1 to 100 Hz) , (Partners 3 and 5) .

Tests in vibration will be carried out either to carry out more total measurements on a structure (analyzes modal, Partenaire 2) or to create separations on the areas of the bad adhesions for an ultrasonic proof test (Partner 6) . These defects could then be detected quickly with rotary sensors without fluid of coupling (Partner 1 with theoretical support of Partner 7) .

All these studies will be carried out with the aim of constituting an expert system starting from the extraction of representative information of the required defects (Partner 2 in collaboration with the other Partners) .

POST-EDITING N°2

NON DESTRUCTIVE TESTING TECHNIQUES

A wide range of industries uses bonded assemblies but their applications are limited because of a lack of non destructive testing techniques for detecting defects which could limit the reliability and the life of the structure. These defects are above all problems of adhesion between the adhesive and the part and problems of cohesion (quality of the adhesive) which must be detected and evaluated by a non destructive method. The aim of this project is to gather European skills in order to make a decisive step forward in the control of these structures (and of the interfaces in general) by developing new ultrasonic techniques which permit an improvement in speed and capacities of detection. These studies form a comprehensive programme of the best techniques that can be considered in this field. The interconnection between these various laboratories will make it possible to compare and supplement techniques which will then be evaluated in a round robin test. The repercussions are significant in the aerospace, car and even electronics industry.

If a large number of industries are tempted to use the bonded assemblies, they generally come up against the problem of the quality assurance of the end product. This quality is based on the control of the production methods but also on a non destructive testing able to detect defects which could limit the structure life.

These defects of the cohesive type (porosity, faulty curing) or adhesive type (lack of contact or contact without adhesion) can occur during production and worsen whilst in service. The most adapted method for detection of these defects are above all ultrasonic techniques. All the partners in this project have already a serious experience on these techniques, either considering research development or use of devices already marketed by partner 1 and 3.

Gathering European skills will permit a comparative evaluation of the new innovative techniques, using the reflections of the different partners as well as studies currently carried out in the USA.

All the partners in this project have already a serious technical experience on the subject and we can ensure that this programme will permit to obtain completely satisfactory results. Moreover, Partners 1, 2, 3 have a great experience in multipartner programmes, which is an additional asset for the success of this project.

- Reduction of control times (factor 5)

- To deal with the foreign competition by proposing more reliable, better conceived products in the aerospace, car or electronics industry.

- Increase in reliability by the use of a printout system (automatic or semiautomatic) connected to an expert system limiting human interpretation.

- Better knowledge of the behaviours of adhesives and adhesive bonded joints, leading to a more rational use of this joining method.

- Weight saving because of the possibility to use local reinforcements where stresses are high.

6.1 General presentation

The modern industry turns more and more towards the use of materials made of layers or successive protections, which require to set up non destructive control methods in order to ensure quality of the interfaces and of the finished product. This is particularly true for bonded assemblies which have many advantages (weight saving, better stress distribution) and which allow structure designs or positionings impossible to achieve by welding or riveting.

However the intensive use of these assemblies encounters problems precisely connected with the quality assurance of the product manufactured.

6.2 Scientific contents of the project

To achieve the objectives described in paragraph 2 and to solve the problems mentioned above, we will develop new ultrasonic techniques.

Fundamental aspects will be approached while studying the properties of the adhesives and of the assemblies aiming at meeting the needs of the design offices for the dimensioning and the understanding of the phenomena of degradation on a micrometric scale (Partner 1 and 4).

Studies will be carried out for detecting adhesion defects by using shear or Lamb waves on a wide frequency range (from 1 to 100 Hz), (Partner 3 and 5).

Vibration tests will be carried out either to perform more global measurements on a structure (modal analysis Partner 2) or to create debondings on the areas of faulty bonding adhesions for a proof ultrasonic test (Partner 6). These defects could then be quickly detected with wheel sensors without couplant fluid (Partner 1 with theoretical support of Partner 7).

All the studies carried out will aim at constituting an expert system from the extraction of data representative of the defects investigated (Partner 2, in connection with other Partners).

TEXT N° 3

"Benefits of Computer Assisted Translation for the Heads of Information Centers (Background paper)

I LANGUAGE PAIR :

French into English

II GLOSSARIES SELECTED :

- Computers / Data processing
- Political science
- Aviation and Space

III SUCCESSIVE OPERATIONS OF THE MACHINE TRANSLATION PROCESS :

- Optical reader --> Time : 2 mn
- Sending a text for translation --> Time : 2 mn
- Post editing / Minimum --> Time : 5 mn
  / Refined --> Time : 10 mn
- Number of words translated --> 379

IV ANALYSIS OF THE ROUGH MACHINE TRANSLATION :

a/ Terminology

Source text	Rough machine translation	Human translation
Intérêt (de qch pour qqn)	Interest of sth for sb	Benefits of sth for sb
Responsables (de centres d'information)	Persons responsible (for centers of information)	Heads of information centers
Applications	Implementations	Applications
Société	Society	Company
Disposition	Provision	Layout
Présenter	To forward	To present

b/ Grammar

Source text	Rough machine translation	Human translation
Textes à traduire	Texts for translation	Texts to be translated

c/ Defective analysis / prepositions

Source text	Rough machine translation	Human translation
Besoins de qqn	Needs for sb	Needs of sb
Attacher de l'importance à.. ni à ...	To attach importance to ... nor with ...	To attach importance to ... nor to ...
Qualité de qch	Quality for sth	Quality of sth
Extérieures à	External at	External to
Apporter qch à ... et à ...	To bring sth to ... and with	To bring sth to ... and to ...
Des exemples ... seront ...	Of the examples will be ...	Examples will be ...
En développement	In development	Under development

d/ Word order

Source text	Rough machine translation	Human translation
Traduction assistée par ordinateur	Translation Computer-assisted	Computer Assisted Translation
N'ont en général pas besoin de...	Do not need in general to	Generally do not need to
Sera normalement suffisante	Will be normally sufficient	Will normally be sufficient
Problèmes techniques et humains	Problems technical and human	Technical and human problems

V FINAL REMARKS

Summaries of voluminous documents or conferences may provide a very useful first approach to a new text. A machine translation system is thus a high-performance tool enabling for example librarians to rapidly know, in their own language, the broad content of a document thus making it much easier for them to file and classify a large amount of texts.

SOURCE TEXT N°3

Intérêt que peut présenter la Traduction Assistée par Ordinateur pour les responsables de centres d'information
(Document de travail provisoire)

Résumé.

Dans cet exposé, les deux applications de Traduction Assistée par Ordinateur seront abordées : la Traduction Assistée par Ordinateur qui a pour but de produire des textes destinés à être diffusés à l'extérieur d'une société et la Traduction Assistee par Ordinateur qui vise à rassembler des informations pour des applications internes.

La dernière application exige des lexiques considérables, couvrant un large éventail de textes et de domaines techniques mais elle n'accorde pas d'importance à la disposition ni à la présentation des informations. Dans ce cas, les textes traduits n'ont en général pas besoin d'être corrigés car la traduction brute sera normalement suffisante pour que les utilisateurs finaux aient une idée approximative du contenu des textes. La vitesse, cependant, est importante.

Si les systèmes de Traduction Assistée par Ordinateur sont utilisés pour produire des publications destinées à des personnes extérieures à la société, d'autres critères doivent être pris en considération, c'est-à-dire :

- Facilité d'importation des textes à partir de divers serveurs,
- Conservation de la présentation et de la disposition des informations,
- Possibilité pour l'utilisateur final de mettre à jour des lexiques et d'influencer la traduction,
- Très bonne qualité de la traduction,
- Outils à disposition pour la correction de la traduction brute,
- Vitesse de traduction,
- Capacité du système à intégrer des développements ultérieurs.

La conférence intitulée "l'intérêt de la Traduction Assistée par Ordinateur pour les responsables de centres d'information et pour les utilisateurs finaux" a pour but de montrer l'intérêt que la Traduction Assistée par Ordinateur peut apporter non seulement au responsable d'un centre d'information, mais également à l'utilisateur final. Après avoir défini les systèmes existants, la nature des textes à traduire, les problèmes techniques et humains liés à l'utilisation des systèmes et les besoins des utilisateurs finaux (qualité des traductions, connaissance de l'information dans la langue maternelle...), des exemples d'applications en cours ou en développement seront présentés. Ces diverses applications permettront de mettre en évidence les avantages de ces systèmes pour les centres d'information et de proposer des solutions au bénéfice de l'utilisateur final.

ROUGH MACHINE TRANSLATION N°3

```
MOTS TRAITES :      379 COMPTE AVANT :    256101 - COMPTE APRES :
    256480-
0100100000P502    FE     YPA TG=3P4DEBUG=S SYS=UCDATE=19 10 89
09H41    0002668
```

Interest of the Translation Computer-assisted for the persons responsible for centers of information (provisional working document)

Summary.

In this statement, the two following implementations of Translation Computer-assisted will be approached : The Translation Computer-assisted which is intended to produce texts intended to be diffused outside a society and the Translation Computer-assisted which aims at gathering information for internal implementations.

The last implementation requires considerable dictionaries, covering a broad range of texts and technical fields but it does not attach importance to the provision nor with the presentation of information. In this case, the translated texts do not need in general to be corrected for the raw translation will be normally sufficient so that the end-users have an approximate idea of the contents of the texts. Speed, however, is significant.

If the translation systems Computer-assisted are used to produce publications intended to persons external at society, other criteria must be taken into account, i.e. :

- Facility of importation of the texts starting from various hosts,
- Conservation of the presentation and the provision of information,
- Possibility for the end-user of updating dictionaries and of influencing the translation,
- The Very good quality of the translation,
- Tools at diposal for the correction of the raw translation,
- Speed of translation,
- Capacity of the system to integrate later developments.

The conference entitled " the interest of the Translation Computer-assisted for the persons responsible for centers of information and for the end-users " is intended to show the interest that the Translation Computer-assisted can bring not only to the person responsible for a center of information but also with the end-user. After having defined the existing systems, the nature of the texts for translation, problems technical and human connected with the use of the systems and the needs for the end-users (quality for the translations, knowledge of information in the mother tongue. .), of the examples of implementations in progress or in development will be forwarded. These various implementations will allow to highlight the advantages of these systems for th centers of information and to propose solutions for the benefit of the end-user.

POST-EDITING N°3

Benefits of Computer Assisted Translation for the Heads of Information Centers
(Background paper)

Abstract:

In this paper, the two following applications of Computer Assisted Translation will be dealt with : the Computer Assisted Translation intended to produce texts to be dispatched outside a company and the Computer Assisted Translation which aims at gathering information for internal applications.

The latter application requires extensive lexicons, covering a wide range of texts and technical fields but need not be concerned with the layout nor with the presentation of information. In this case, the translated texts generally do not need to be corrected, for the rough translation will normally be sufficient for the end users to have an approximate idea of the content of the texts. Speed, however, is significant.

If Computer Assisted Translation systems are used to produce publications intended to third parties, other criteria must be taken into account, i.e. :

- Ease of text import from various host systems,
- Preservation of the presentation and the layout of information,
- Possiblity for the end-user of updating lexicons and of influencing the translation,
- A very high translation quality,
- Tools at disposal for the correction of the rough translation,
- Speed of translation,
- Capacity of the system to incorporate future developments.

The conference entitled "the benefits of the Computer Assisted Translation for the heads of information centers and for the end-users" is intended to show the interest that Computer Assisted Translation can bring, not only to the head of an information center, but also to the end-user. After having defined the existing systems, the nature of the texts to be translated, the technical and human problems connected with the use of the systems and the needs of end-users (quality of the translations, information knowledge in the mother tongue...), examples of on-going applications and systems under development will be presented. These various applications will make it possible to highlight the advantages of these systems for information centers and to propose solutions for the benefit of the end-user.

Bibliography

The bibliography that follows was prepared by CEDOCAR (the French Defence Documentation Centre) and has been approved by the Lecture Series Director, Mr. O. Lavroff of Aerospatiale, France.

To save space the field identifiers for each item have been omitted. This should cause little problem but to help understanding, the fourth item is reprinted below with the identifiers in English and in French. Note that different types of document and documents from different sources do of course have somewhat different fields.

Accession Number/NUMERO SIGNALEMENT	C-89-003619
French Title/TITRE FRANCAIS	Les projets actuels du "Groupe d'études pour la traduction automatique" sur, et à propos de, la **traduction** **assistée** par ordinateur (TAO).
English Title/TITRE ANGLAIS	Current projects at GETA on or about machine translation.
Author(s)/AUTEUR(S)	BOITET C.
Affiliation/AFFILIATION	GETA, Grenoble, FR
Type of Document/TYPE DE DOCUMENT	Publication en série
Language Code/CODE LANGUE	ENG
Country Code/CODE PAYS D'ORIGINE	FR
Periodical Title/TITRE DU PERIODIQUE	Science et Défense (FR)
Issue No., etc/SOURCE	NO 2; 28 p.; 35 Ref.; DP. 1988/05
CODEN/CODEN	SCDF2X
Location (in CEDOCAR)/GISEMENT	05; M6039-3
Abstract/RESUME	Le GETA poursuit ses recherches fondamentales et appliquées sur la **traduction** **assistée** par **machine**, avec construction de systèmes expérimentaux et étude des techniques associées à l'intelligence artificielle et à la linguistique moderne.De nouvelles idées sont examinées pour améliorer l'approche par transfert désormais classique, de nouveaux langages sont ajoutés, des parties des théories modernes linguistiques sont incorporées dans les grammaires et des environnements de logiciels plus performants sont construits pour les langages spécialisés à base de règles.L'objectifs primordial est la construction d'une station de travail pour les linguistes, les lexicographes et les utilisateurs finals.
Origin of abstract/SIGNATURE ANALYSTE	INFO/CR
Classification Code/CODE CLASSIFICATION	05 07
Descriptors/DESCRIPTEUR(S)	Traduction machine*;Linguistique automatisée;Traduction automatique;Intelligence artificielle

NB the Classification Code uses the COSATI Field and Group structure.

the words in the abstract that are bracketed by ** and ** are the terms used for the search that produced this bibliography.

BM-89-000473
GWAI 87.Onzième congrès germanique sur l intelligence artificielle.
GWAI 87.11th German worshop on artificial intelligence.
Geseke (DE)
1987/09/28-1987/10/02
MORIK K.
Congrès
MUL
ZZ
Springer (Berlin)
405 p.; ND. Ref.; ND. Fig.; DP. 1987
3-540-18388-4
03; 412-152
Langages naturels.Traduction automatique.Analyse grammaticale et analyse syntaxique.Représentation des connaissances.Systèmes experts.Systèmes déductifs.
INFO/GD
06 04
Intelligence artificielle*,Langage naturel.Traduction automatique; Traduction machine;Analyse syntaxique:Grammaire;Représentation connaissance;Système expert

C-89-005008
Coling 88 12ème congrès international sur la linguistique informatisée.
Coling BUDAPEST 12th international conference on computational linguistics.
BUDAPEST (HU)
1988
JOHN VAN NEUMANN society for computing sciences (HU)
Congrès
ENG
HU
John Van Neuman society for computer sciences (Budapest)
845 p.; NB Ref.; NB Fig.; DP. 1988
9-638-43156-3
05; M 3242
Traitement du langage naturel.**Traduction** **machine**, acquisition des informations sémantiques, grammaires.Traitement linguistique pour la reconnaissance et la compréhension de la parole.Développement de grammaires de langage naturel.
INFO/GD
05 07
Linguistique automatisée*;Traduction automatique;Traduction machine;Langage naturel;Sémantique;Grammaire

C-89-003620
Représentation et information d'équipements de traduction pour une interprétation automatique de textes parlés
Representation and computation fo units of translation for machine interpretation of spoken texts.
BOITET C.
Geta, grenoble (fr)
Publication en série
ENG
FR
Sciences et Défense (R)
NO 2; 41 p.; 122 Ref.; 11 Fig.; DP. 1988/05
SCDF2X
05; M 6039-3
Comme la traduction automatique, l'interprétation assistée par ordinateur est une entreprise industrielle, impliquant synthèse et reconnaissance de parole elle est néanmoins plus complexe.L'architecture proposée dans cet article peut aussi sembler complexe, elle est pourtant maintenue aussi simple que possible grâce à une structure de donnée d'un type unique compatible avec toutes les méthodes algorithmiques appliquées avec succès en reconnaissance de parole, traduction automatique et compréhension de langage, et compatible également avec les futurs développements des réseaux neuronaux.
INFO/CR
05 07; 06 04
Traduction machine*;Intelligence artificielle*;Traduction automatique;Reconnaissance parole;Procedure arborescente;Semantique
Reseau neuronal;Synthese parole;Analyse syntaxique

C-89-003619
Les projets actuels du "Groupe d'études pour la traduction automatique" sur, et à propos de, la **traduction** **assistée** par ordinateur (TAO).
Current projects at GETA on or about machine translation.
BOITET C.
GETA, Grenoble, FR
Publication en série
ENG
FR
Science et Défense (FR)
NO 2; 28 p.; 35 Ref.; DP. 1988/05
SCDF2X
05; M6039-3
Le GETA poursuit ses recherches fondamentales et appliquées sur la **traduction** **assistée** par **machine**, avec construction de systèmes expérimentaux et étude des techniques associées à l'intelligence artificielle et à la linguistique moderne.De nouvelles idées sont examinées pour améliorer l'approche par transfert désormais classique, de nouveaux langages sont ajoutés, des parties des théories modernes linguistiques sont incorporées dans les grammaires et des environnements de logiciels plus performants sont construits pour les langages spécialisés à base de règles.L'objectifs primordial est la construction d'une station de travail pour les linguistes, les lexicographes et les utilisateurs finals.
INFO/CR
05 07
Traduction machine*;Linguistique automatisée;Traduction automatique;Intelligence artificielle

C-88-014117
Barrières aux transferts d informations et quelques approches vers leurs réductions.
Barriers to information transfer and approaches towards their reduction.
Washington (US)
1987/09/23-1987/09/24
Agard (fr)
Congrès
ENG
FR
AGARD (Neuilly/Seine)
VOL CP 430; 116 p.; ND. Ref.; DP. 1988/03
9-283-50449-6
02; AGARD-CP-430
Barrières aus transferts d'informations entre les pays membres de la Communauté Aérospatiale de Recherche et de Défense.Définition de l'amplitude du problème.Identification des barrières (économiques, linguistiques et culturelles).Relations hommes machines.Traduction automatique.
INFO/GD
09 04
Transfert information*;Barrière;Traduction machine;Recherche défense
Echange information*;Echange international*;Barrière linguistique, Recherche aérospatiale

C-88-009085
Traduction avec accès mémoire direct.
Direct memory acces translation.
10th International Joint Conference on Artificial Intelligence (IJCAI 87).
Milan, IT
1987/08/23-1987/08/28
TOMABECHI H.
Carnegie Mellow Univ., Pittsburgh, US
Mémoire Congrès
ENG
US
Morgan Kaufmann (Los Altos)
VOL 2/2; pp. 722-725; 16 Ref.; 1 Fig.; DP. 1987
0-934-61343-5
05; M 5234
Présentation d'une théorie dans laquelle la traduction est considérée comme partie intégrante du traitement cognitif.Dans ce paradigme, la compréhension en langage source est une reconnaissance des entrées, en termes de connaissances existant en mémoire, suivie d'une intégration de ces entrées dans la mémoire.La traduction étant effectuée avec accès direct au réseau de mémoire d'autres processus cognitifs (inférence, par exemple) peuvent participer dynamiquement à cette traduction (création de nouveaux concepts;apprentissage d'un nouveau vocabulaire).
INFO/CR
05 07; 06 04
Traduction machine*;Machine apprentissage*
Accès mémoire;Théorie de la connaissance;Analyse lexicale; Traitement parallèle;Règle inférence;Partage mémoire;langage naturel

C-88-009084
Architecture d'analyseur syntaxique universel pour ***raduction** par **machine** intelligente.
The universal parser architecture for knowledge-based machine translation.
10th International Joint Conference on Artificial Intelligence (IJCAI 87).
Milan, IT
1987/08/23-1987/08/28
TOMITA M.; CARBONELL J. G.
Carnegie Mellon Univ., Pittsburgh, US;Carnegie Mellon Univ., Pittsburgh, US
Mémoire Congrès
ENG
US
Morgan Kaufmann (Los Altos)
VOL 2/2; pp. 718-721; 17 Ref.; 2 Fig.; DP. 1987
0-934-61343-5
05; M 5234
Une **traduction** par **machine** doit être sémantiquement précise, linguistiquement correcte, interactive et extensible à plusieurs langues et domaines.Une architecture d'analyseur universel s'efforce d'atteindre l'ensemble de ces objectifs.Des bases de connaissances linguistiques (syntaxe, sémantique, lexique, pragmatisme) codées sous des formes appropriées (grammaire) sont modifiées et précompilées en vue de l'analyse syntaxique et de la génération des textes.Les premiers résultats de traductions bidirectionnelles anglais et japonais snt encourageants et démontrent la faisabilité théorique de l'approche.
INFO/CR
05 07; 09 02
Traduction machine*;Architecture calculateur*;Sémantique; Linguistique automatisée;Grammaire
Analyse syntaxique*;Base de connaissance

C-88-008796
Analyse syntaxique structurale modifié pour systèmes de compréhension de langage.
Modified caseframe parsing for speech understanding systems.
10th International Joint Conference on Artificial Intelligence (IJCAI 87).
Milan, IT
1987/08/23-1987/08/28
POESIO M.; RULLENT C.
CSELT, Torino, IT;CSELT, Torino, IT

Mémoire Congrès
ENG
IT
Morgan Kaufmann (Los Altos)
VOL 2/2; pp. 622-625; 12 Ref.; 6 Fig.; DP. 1987
0-934-61343-5
05; M 5234
Proposition d'une stratégie d'analyse de formes syntaxiques, pour systèmes de compréhension de parole, différant des stratégies classiques par aux moins deux aspects : l'analyse ne repose pas uniquement sur un processus descendant et les formes syntaxiques sont amalgamées à une connaissance syntaxique avant d'être utilisées.Cette stratégie peut être exécutée comme un processus d'inférence.
INFO/CR
17 02; 05 07
Reconnaissance parole*;Sémantique*;Linguistique automatisée; Traduction machine
Analyse syntaxique*;Esprit programme;Forme syntactique;Règle inférence,Langage naturel

C-88-008430
Constructtion d'interfaces en langage naturel pour systèmes experts basés règles.
Bulding natural language interfaces for rule-based expert systems.
10th International Joint Conference on Artificial Intelligence (IJCAI 87).
Milan, IT
1987/08/23-1987/08/28
MOERDLER G. D.; MCKEOWN K. R.; ENSOR J. R.
Columbia Univ., New York, US;Columbia Univ., New York, US;AT and T Bell Labs., Molmdel, US
Mémoire Congrès
ENG
US
Morgan Kaufmann (Los Altos)
VOL 2/2; pp. 682-687; 25 Ref.; 5 Fig., DP. 1987
0-934-61343-5
05; M 5234
Etude d'une sémantique pour traduction de phrases en langage naturel, en informations factuelles pour un système expert sous-jacent, en remplacement de l'interface à menu plus conventionnelle utilisée pour collecter les données fournies par l'utilisateur.Description de deux problèmes rencontrés dans la construction de ce nouveau type d'interfaces pour systèmes experts : le traitement sémantique des phrases de l'utilisateur et la conception d'un interprète, pour le système expert, utilisant efficacement les données factuelles fournies par l'utilisateur.
INFO/CR
09 02; 05 07
Système expert*;Traduction machine*;Relation homme machine; Sémantique;Acquisition donnée;Interprétation
Langage naturel*;Base donnée factuelle;Programme traducteur;Moteur d inférence;Base de connaissance

C-88-008328
Compréhension de spécifications de systèmes, écrites en langage naturel.
Understanding system specifications written in natural langage.
10th International Joint Conference on Artificial Intelligence (IJCAI 87).
Milan, IT
1987/08/23-1987/08/28
GRANACKI J. J.; PARKER A. C.; ARENS Y.
Univ.of Southem California, Los Angeles, US;Univ.of Southem California, Los Angeles, US;Univ.of Southem California, Los Angeles, US
Mémoire Congrès
ENG
US
Morgan Kaufmann (Los Altos)
VOL 2/2; pp. 688-691; 12 Ref.; 2 Fig.; 1 Tabl.; DP. 1987
0-934-61343-5
05; M 5234
Description de recherches sur la compréhension de spécifications systémiques écrites en langage naturel Ces recherches comportent la mise en oeuvre de l'interface PHRAN-SPAN ("Phrasal Analyser-Specification Analysis"), pour la spécification du comportement abstrait de systèmes numériques dans un texte anglais limité, avec le système ADAM ("Advanced Design Auto Mation") de l'University of Southem California.
INFO/CR
05 07
Sémantique*;Conception assistée par calculateur;Traduction machine; Spécification
Langage naturel*;Analyse syntaxique;Etude conception système; Structure donnée

C-88-008327
Représentation et interprétation de déterminants dans un langage naturel.
Representation and interpretation of determiners in natural language.
10th International Joint Conference on Artifical Intelligence (IJCAI 87).
Milan, IT
1987/08/23-1987/08/28
DI EUGENIO B.; LESMO L.
Univ., Torino, IT;Univ., Torino, IT
Mémoire Congrès
ENG
IT
Morgan Kaufmann (Los Altos)
VOL 2/2; pp. 648-654; 22 Ref.; 3 Fig.; DP. 1987
0-934-61343-5
05; M 5234
Proposition d'un formalisme de représentation sémantique susceptible de traiter les problèmes de référence c'est à dire d'interpréter des séquences de mots (notamment celles débutant par un article).Ce formalisme suit une approche de réseau sémantique et utilise différents plans de représentation (sémantique. contenu, référence) ainsi que certaines structures particulières, denommées espaces d'ambiguité, qui dissimulent les ambiguités et restent neutres vis à vis des diverses interprétations possibles jusqu'à ce que les ambiguités soient levées.
INFO/CR
05 07
Sémantique*;Interprétation;Linguistique automatisée;Phrase grammaire;Traduction machine
Langage naturel*;Base de connaissance;Fonction ambiguité;Grammaire syntactique

C-88-005301
Colloque sur les techniques d'évaluation pour conception de systèmes interactifs (2e).
Colloquium on evaluation techniques for interactive system design : II.
London, GB
1987/10/02
IEE Comput.and Ctrl.Division (GB)
Congrès
ENG
GB
IEE Colloquium Digests (GB)
IEE, London
NO 1987/78; 26 p.; 2 Ref.; 1 Fig.; 3 Tabl.; 7 résumés; DP. 1987
IEECDB
05; Me 131-4
Problèmes pratiques rencontrés dans l'analyse de données d'interactions temps réel homme-machine.Utilisation d'équipements vidéo dans le processus de conception.Analyse de questionnaires sur la satisfaction des utilisateurs de l'informatique.Application de techniques statistiques au processus de conception en vue de minimiser l'évaluation expérimentale.Etude des avantages d'une planification en profondeur rigoureuse dans le domaine de la recherche appliquée.Description d'une méthodologie de conception de systèmes de traitement de données.Présentation d'un outil de conception pour interfaces graphiques interactions.
INFO/CR
05 08; 09 02
Relation homme machine*;Affichage graphique interactif*;Interface; Recherche appliquée;Planification projet;Assurance qualité; Traduction machine;Psychométrie;Conception assistée par calculateur;Evaluation performance
Système interactif*;Etude conception système

C-88-F01168
Technologie nordique de pointe
Fonds de l'Industrie Nordique (NO)
Ouvrage
FRE
NO
Fonds de l'Industrie Nordique, Oslo
39 p.; quelq. Fig.; nombr. Phot.; DP. 1987
05; M 260/824 P
Les télécommunications par satellite.La coordination de la fabrication de circuits VLSI.La réutilisation systématique de modules de programme.La **traduction** **assistée** par ordinateur.La synthèse de la parole.Le traitement des images.Le transfert automatique sur bande.Etude des enzymes.Biotechnologie et protéines.Les systèmes d'aide à la navigation maritime par gestion informatisée interactive.Propriétés mécaniques des alliages soumis à solidification et extrusion rapides.Nouveaux champs d'application pour les composites époxy-fibres de carbone.
INFO/AN
14 06
Recherche développement*,Danemark;Finlande;Islande;Norvège.Suède; Traduction machine;Traitement image;Langage programmation; Transfert information,Navigation maritime;Télécommunication par satellite.
Coopération scientifique*;Développement technologique*;EUREKA projet;Intégration très grande échelle;Pays nordique

C-88-F00089
Intelligence artificielle et systèmes experts.
Convention informatique.L'informatique : du discours à la méthode.
Paris (FR)
1986/09/15-1986/09/19
Direction des Industries Electroniques et de l'Informatique (FR)
Mémoire Congrès
FRE
FR
DIELI, Paris
VOL A; pp. 45-60; nombr. Fig.; DP. 1986
2-902-57421-5
05; M 5743
L'acquisition de la maitrise industrielle en systèmes experts : mise en oeuvre de systèmes experts dans une entreprise.Une expérience réussie de transfert technologique entre la recherche et une application opérationnelle : rôle et action de différents intervenants du monde de la recherche et de l'entreprise à l'occasion du développement d'un Système Expert en diagnostic de pannes.Le traitement automatique du langage : l'avenir de la TAO (**Traduction** **Assistée** par Ordinateur), des interfaces de dialogue, d'aide à la rédaction, machines à dicter et de compréhension de texte.

INFO/VT
09 02
Informatique*;Système expert*;Intelligence artificielle*;Traduction machine;Diagnostic;Panne;Traitement automatique données;Texte;Reconnaissance parole;Rédaction;Etude développement Traitement texte

C-87-013628
Intelligence artificielle : outils, méthodes et applications.
Convention informatique.L'informatique : du discours à la méthode.
Paris (FR)
1986/09/15-1986/09/19
Direction des Industries Electroniques et de l'Informatique (FR)
Mémoire Congrès
MUL
FR
DIELI, Paris
VOL A; pp. 124-195; nombr. Ref., nombr. Fig., anglais, français. DP. 1986
2-902-57421-5
05; M 5743
Représentation des connaissances (langages et méthodes) : architecture d'intelligence service, introduction à OPS5, apprentissage de concepts, GURU : un outil de développement de systèmes experts dans le monde de la gestion.Architecture de cinquième génération : architecture des machines USP, méthodes de programmation parallèle d'une architecture MIMD, DDC : Delta Driven Computer, une architecture pour l'intelligence artificielle.Des systèmes experts aux systèmes à base de connaissance . l'intelligence artificielle face aux techniques informatiques, application des systèmes experts à la conception des bases de données, l'apport des techniques de l'I.A.dans un environnement bureautique.Applications des aspects, reconnaissance des formes, langue naturelle de l'IA : le projet de **traduction** **assistée** par ordinateur (TAO) langage naturel et recherche documentaire, génération automatique de textes en langues naturelles, la reconnaissance de la parole.
INFO/VT
09 02; 06 04
Informatique*;Intelligence artificielle*;Base donnée;Système expert;Architecture calculateur;Bureautique;Reconnaissance forme;Traduction machine;Reconnaissance parole;Recherche documentaire Architecture système;LISP langage programmation;Base de connaissance;Représentation connaissance;Ordinateur cinquième génération

C-87-011588
Langage naturel.
Natural language.
9th International Joint Conference on Artificial Intelligence.
Los Angeles (US)
1985/08/18-1985/08/23
Ijcai/aaai/ucla (us)
Mémoire Congrès
ENG
US
Western Periodicals Co., North Hollywood
VOL 2/2; pp. 749-885; nbr. Ref.; nbr. Fig.; DP. 1985
0-934-61302-8
05; M.5234
Près de trente mémoires consacrés aux langages naturels traitent: d'analyses syntaxique et syntagmatique, de grammaires, d'interactions entre syntaxe et sémantique, de génération de langage naturel et de traduction automatique, de processeurs de langage d'analyse de langue chinoise, de structures de discours et de résumés de textes, de conversation homme-machine, d'accès à des bases de données intelligentes, de système d'information intelligent de traitement d'inférences en ligne directe, d'analyse de conjonctions à l'aide du langage Prolog, de reconnaissance automatique de parole.
INFO/CR
05 07; 09 02
Linguistique automatisée*;Langage programmation*;Relation homme machine;Sémantique;Grammaire;Intelligence artificielle;Langage indépendant contexte;Traduction machine;Langue chinoise;Système information;Programmation dynamique
Langage naturel*;Théorie langage;Analyse syntaxique;Reconnaissance automatique parole;Machine turing;Automate fini;Traitement texte

C-87-004473
Méthode d'analyse syntaxique du langage naturel par une procédure de filtrage.
A parsing method of natural language by filtering procedure.
SAKAKI H.; HASHIMOTO K.; SUZUKI M.; NOGAITO I.; TANAKA T.
Kokusai Denslim Denwa Co., Tokyo, JP;Kokusai Denslim Denwa Co., Tokyo, JP;Kokusai Denslim Denwa Co., Tokyo, JP;Advanced Telecom.Res.Inst.Int.; Osaka-shi, JP;Software Consult.Co., Tokyo, JP
Publication en série
ENG
JP
Transactions of the Institute of Electronics and Communication Engineers in Japan (JP)
VOL E 69; NO 10; pp. 1114-1124; 6 Ref.; 8 Fig.; DP. 1986/10
TIEEDU
0387-236X
05; P 1725
Présentation d'une méthode d'analyse syntaxique, basée sur une extension de "LINGOL", comportant deux étapes : une décomposition arborescente sans limitation, suivie d'une élimination des branches incorrectes grâce à un filtre approprié utilisant des notions d'"arbre interdit" et d'"arbre privilégié".Description du traitement des sommets logiques "OU" permettant d'exprimer plusieurs arbres sur un arbre unique.Etude de l'application pratique de cette méthode d'analyse dans le système "KATE" de traduction automatique d'anglais en japonais.
INFO/CR
05 07
Traduction machine*
Analyse syntaxique*;Langage naturel*;Procédure arborescente;Traduction dirigée syntaxe;Analyse assistée par calculateur;Filtre rejecteur

BM-87-000314
Applications de microprocesseurs à l'intelligence artificielle.
Microprocessor applications in artificiel intelligence.
12 EUROMICRO symposium on microarchitectures, developments and applications.
Venise, IT
1986/09/15-1986/09/18
EUROMICRO Assoc.for microprocessing and microprogramming (NL)
Mémoire Congrès
ENG
NL
Elsevier Science Publishers B.V.(Amsterdam)
pp. 69-95; NB Ref.; NB Fig.; 3 mémoires; DP. 1986
0-444-70096-X
05; M 5872
Proposition d'un préprocesseur intégré pour l'exécution de programmes en PROLOG basée sur un ensemble d'instructions de type Warren.Etude d'une nouvelle structure du calcul pour la traduction du langage LISP applicable aux microcalculateurs.Présentation d'une architecture de logiciel pour estimateur de position de robot mobile dans un environnement limité, bureautique ou domestique.
INFO/CR
06 04; 09 02
Intelligence artificielle*;Microprocesseur*;Séquence instruction;Architecture calculateur;Microinstruction;Traduction machine;Compilateur;Robot
Préprocesseur;PROLOG langage programmation;LISP langage programmation

BM-87-000105
Reconnaissance de forme.
Pattern recognition.
Applications of artificial intelligence III.
Orlando (US)
1986/04/01-1986/04/03
The International Society for Optical Engineering
Mémoire Congrès
ENG
ZZ
Proceedings of SPIE (US)
SPIE, Bellingham
VOL 635; pp. 439-496; nombr. Ref.; nombr. Fig.; nombr. Tabl.; 6 communications; DP. 1986
SPIECJ
0-892-52670-X
05; Me 10628
Présentation d'une méthode de reconnaissance automatique des changements de primitives dans les signaux de commande numérique des machines.Description d'un système fiable de traduction automatique par calculateur.Application d'un système expert pour les diagnostics de la médecine chinoise traditionnelle.Reconnaissance de caractères appartenant à plusieurs polices en utilisant des techniques d'apprentissage.
INFO/MD
06 04
Reconnaissance forme*;Reconnaissance caractère;Intelligence artificielle;Système commande;Commande numérique;Traduction machine Diagnostic clinique;Système expert

C-86-012650
L'intelligence artificielle.
Artificial intelligence.
SATO S.; SUGIMOTO M.
Publication en série
ENG
ZZ
Fujitsu Scientific and technical Journal (JP)
VOL 22; NO 3; pp. 139-181; 66 Ref.; 36 Fig.; 4 Tabl.; DP. 1986
FUSTA4
0016-2523
05; P1708
Etat de l'art de la recherche et du développement en matière d'intelligence artificielle dans la Société Fujitsu qui participe actuellement à un projet d'ordinateurs de la cinquième génération.
INFO/TT
09 02; 05 04
Système expert*;Intelligence artificielle*;Traduction machine*;Traitement information;Reconnaissance parole;Langage programmation;Mémoire virtuelle calculateur;Conception assistée par calculateur

C-86-010787
Dixième congrès international sur la linguistique automatisée.
10th international conference on computational linguistics.22nd annual meeting of the Association for computational linguistics.
Stanford, US
1984/07/02-1984/07/06
Association for Computational Linguistics (US)
Congrès
ENG
US
Acl (us)

561 p.; No. Ref.; No. Fig.; DP. 1984
05; M 3242
Traduction par **machine**.Analyse grammaticale.Analyse sémantique.Interfaces en langage naturel.Analyse syntaxique.Analyse syntactique.Lexicographie.Compréhension automatique des textes.
INFO/GD
05 07
Traduction machine*;Linguistique automatisée*
Grammaire syntactique;Analyse syntaxique;Langage naturel;
Lexicographie;Discours;Base de connaissance

-86-010286
Un langage simple d'application : le mini M-L.
A simple applicative language : mini M-L.
CLEMENT D.; DESPEYROUX J.; DESPEYROUX T.; KAHN G.
SEMA (FR);INRIA Sophia Antipolis (FR);INRIA Sophia Antipolis (FR);INRIA Sophia Antipolis (FR)
INRIA, Le Chesnay
Rapport
ENG
FR
529
Rapport de recherche; 15 p.; 15 Ref.; 12 Fig.; DP. 1986/05
0249-6399
05; M 5208-4
Description formelle de la partie essentielle du langage ML en Sémantique Naturelle.Les sémantiques statique et dynamique sont traitées ainsi que la **traduction** vers une **machine** abstraite.Cette description a fait l'objet de vérifications sur ordinateur et nous expliquons pourquoi ces vérifications sont possibles.Un certain nombre de propriétés du langage s'expriment aisément dans le contexte de cette méthode et nous les démontrons.
INFO/LP
09 02
Langage programmation*
Calcul lambda*;Formule implicite;Codage numérique

C-86-009146
Numéro spécial consacré au traitement du langage naturel.
Special issue on natural language processing.
Publication en série
ENG
ZZ
Proc.of the IEEE (US)
VOL 74; NO 7; pp. 899-1039; No. Ref.; No. Fig.; No. Tabl.; DP. 1986/07
IEEPAD
0018-9219
05; P 0739
La représentation des connaissances et le traitement du langage naturel.Le langage naturel et les experts artificiels.Les modèles utilisateurs fondés sur le dialogue.La génération du langage.La machine à traduire : perspectives européenne, américaine et japonaise.Evaluation des systèmes de traitement du langage naturel.
INFO/TT
05 07
Langage*;Intelligence artificielle*,Traduction machine*;Système expert*;Traduction;Langage indépendant contexte;Théorie graphe;
Programme calculateur;Système homme machine
Traduction dirigée syntaxe*,Langage naturel*,Base de connaissance*

C-86-004403
Atlas: système de traduction automatique.
Atlas: automatic translation system.
UCHIDA H.; HAYASHI T.; KUSHIMA H.
Fujitsu (JP);Fujitsu (JP);Fujitsu (JP)
Publication en série
ENG
JP
Fujitsu (JP)
VOL 21; NO 3; pp. 317-329; 13 Fig.; DP. 1985/ET
FUSTA4
05; P,1708
Deux machines de traduction chez Fujitsu Atlas I: basée sur la syntaxe et Atlas II basée sur la sémantique.On explique les deux mécanismes de traduction.
INFO/GD
05 07
Traduction machine*;Sémantique;Syntaxe
Traduction dirigée syntaxe

C-86-003758
Préparation d'une base de données de langue anglaise en ligne pour l'information scientifique et technique japonaise.
Preparation of an online English language database for Japanese scientific and technical information.
9th international online information meeting.
Londres, GB
1985/12/03-1985/12/05
MORITA A.; SATO M.; NISHIDA R.
The Japan INF.CENT.of Sci.and Technol., JICST, JP;The Japan INF.CENT.of Sci.and Technol., JICST, JP;The Japan INF.CENT.of Sci.and Technol., JICST, JP
Mémoire Congrès
ENG
JP
Learned Information.Oxford and New Jersey (GB)
pp. 61-67; 2 Fig.; 1 Tabl.; DP. 1985
0-904-93350-4
05; M 5789
Base de données en anglais à l'usage des étrangers créée en 1985 par le Centre d'Information japonaise de science et de technologie.Caractéristiques du fichier JICST, systèmes de conversion du japonais en anglais, avenir du fichier.Le service en ligne de la base de données en anglais sera opérationnel en 1986 sur le serveur JOIS.
INFO/AN
05 01; 05 02
Base de donnée*;Information technique*;Japon;Information scientifique;Langue anglaise;Traduction machine
Système conversationnel interactif;Coopération scientifique

C-86-F00678
5ème congrès.Reconnaissance des formes et intelligence artificielle.2 tomes.
5ème Congrès.Reconnaissance des Formes et Intelligence Artificielle.2 Tomes.
Grenoble (FR)
1985/11/27-1985/11/29
AFCET (FR).Agence de l'Informatique (FR) INRIA (FR)
Congrès
FRE
FR
AFCET, Paris
VOL 1-2; 1283 p.; nbr. Ref.; nbr. Fig.; nbr. Tabl.; DP. 1985
2-903-67711-5
05; M.5836
Historique de l'intelligence artificielle.**Traduction** **assistée** par ordinateur.Les systèmes Cognitifs (Lore, Prolog, le système Machin 2).La segmentation d'images, robotique, dialogue oral homme-machine, langage naturel (notamment , un système expert de traitement des textes écrits).Définition d'une base de données d'images de télédétection.Un système pour la construction d'une hiérarchie d'experts en télédétection.Représentation des connaissances pour un système d'aide intelligent.Vision mobile et stéréo.
INFO/TT
09 02; 06 04
Système expert*;Intelligence artificielle*;Reconnaissance forme*.
Robot;Télédétection;Relation homme machine;Vision stéréoscopique
Robotique*;Traitement texte*;Langage naturel*;Prolog langage programmation

BM-86-000424
L'intelligence artificielle appliquée à la traduction.
AI fine-tunes speech recognition.
GALLAGHER R.
Publication en série
ENG
ZZ
Electronics (US)
VOL 59; NO 20; pp. 24, 25; DP. 1986/05/19
ELECAD
0883-4989
05; P 0213
La firme italienne Olivetti escompte de grands progrès dans les systèmes de traduction par emploi de l'intelligence artificielle.Application au renseignement militaire.
INFO/VZ
06 04; 05 04
Intelligence artificielle*;Renseignement militaire*;Traduction machine*;Italie;Etude conception matériel
Olivetti société*

TIB/A89-82002/XAD
Computer-aided Saarbruecken Translation Service STS.Final report of the MARIS project.
Computergestuetzte Saarbruecker Translationsservice STS.Abschlussbericht des Projekts MARIS.
ZIMMERMANN H. H.; LUCKHARDT H. D.
Universitaet des Saarlandes, Saarbruecken (Germany, F.R.).Fachrichtung Informationswissenschaft.
Bundesministerium fuer Forschung und Technologie, Bonn (Germany, F.R.).
019985016
Report
GER
DE
In German,Veroeffentlichungen der Fachrichtung Informationswissenschaft, With 50 refs; NP. 265; DP. May 89.
U9001
NTIS Prices: PC E07
BMFT 1013209/2
The MARIS project (multilingual application of reference-oriented information systems) has investigated the scientific and technical preconditions for the application of computer-aided and **machine** **translation** in the field of specialized information.MARIS has established a **computer**-**aided** **translation** service for the translation of specialized information from German data bases into English.The report reflects the essential aspects of the project work: integration of **machine** **translation** into a translation service, man-machine interaction at a translator's workbench, development, storage, and use of terminology in computer-aided and **machine** **translation**, multilinguality of specialized information, remaining problems, technical aspects, concrete translations.(orig.).(TIB: FR 2736.) (Copyright (c) 1989 by FIZ.Citation no.89:082002.).
92 04; 88 02
Computer programs*;Machine translation*;Information systems;
Linguistics;Language programming;Dictionaries;Indexes
Documentation;Man machine systems;Personal computers
Foreign technology*;NTISTFFIZ;NTISFNGE;NTISLNGER

PB89-868913/XAD
Chinese and Japanese Language Translation by Computer.January 1975-August 1989 (Citations from the INSPEC: Information Services for the Physics and Engineering Communities Database).
National Technical Information Service, Springfield, VA.
055665000
Report
ENG
US
Rept. for Jan 75-Aug 89; Supersedes PB87-863098; NP. 59; DP. Aug 89.
U8920
NTIS Prices: PC N01/MF N01
This bibliography contains citations concerning research and development of computer hardware and software for the language translation of Chinese and Japanese.Computer technology in character recognition, sentence analysis, text input and output systems, automatic language translation systems for personal computers, and character generation and analysis are discussed.Translation techniques for Chinese-to-Japanese, Chinese-to-English, and Japanese-to-English are presented.Applications in business, utilities management, and library automation are included.(This updated bibliography contains 100 citations, 40 of which are new entries to the previous edition.).
92 04; 88 05
Bibliographies*;Machine translation*;Chinese languages*;Japanese languages*;Automatic language processing*;Input output devices
Computers;Computer systems hardware;Computer systems programs;
Character recognition
Chinese language translation*;Japanese language translation*:
Published Searches;NTISNTISH;NTISNERACD

PB89-867931/XAD
Machine **Translation**: Foreign Language Translation and Natural Language Understanding.January 1970-July 1989 (Citations from the NTIS Database).
National Technical Information Service, Springfield, VA.
055665000
Report
ENG
US
Rept. for Jan 70-Jul 89; Supersedes PB87-866349; NP. 68; DP. Aug 89.
U8920
NTIS Prices: PC N01/MF N01
This bibliography contains citations concerning research and development of machine/mechanical foreign language translation by computer.Topics include syntactic and semantic translations, natural language representation and understanding, knowledge based systems, language manuals for ideographic machines, systran **machine** **translation**, mathematical linguistics and logic, foreign technologies and language translation, processes for question answering, and Chinese lexicography,and romanization.Methods and systems for translations of Russian, German, Chinese, and Japanese to English are presented.(This updated bibliography contains 126 citations, 30 of which are new entries to the previous edition.).
92 04; 88 05
Bibliographies*;Machine translation*;Automatic language processing*;Computational linguistics*;Syntax;Semantics;Artificial intelligence:Translating;English language;Russian language;German language;Chinese language;Japanese language
Foreign languages*;Natural language*;Published Searches;Vocabulary;
NTISNTISN;NTISNERACD

N89-23363/9/XAD
Objectives and Role of the Greek National Documentation Center.
BOUBOUKAS V.; SKOURLAS C.; POULAKAKI E.
National Hellenic Research Foundation, Athens (Greece).
National Aeronautics and Space Administration, Washington, DC.
080563000; NI508359
Report
ENG
GR
In AGARD, the Organisation and Functions of Documentation and Information Centres in Defence and Aerospace Environments 4 p; NP. 4; DP. Mar 89.
S2716
NTIS Prices: (Order as N89-23362/1, PC A06/MF A01)
A brief overview of the Greek information scene is presented.The objectives and the role of the National Documentation Centre are outlined together with some of its activities which proved to function within such an information environment as well as plans for continuity.
88 00; 88 02
Computer programs*;Data management*;Information dissemination*;
Information systems*;Machine translation*;Greece*;Languages;
Alphabets;Bibliographies;Literature;Reports
Foreign technology*;NTISNASAE;NTISFNGR

N89-20661/5/XAD
Barriers to the International Transfer of Information in Aerospace and Defense.
Contained in:AGARD-Cp-430 Accessioned as N88-30458,Presented at the Meeting on Barriers to Information Transfer and Approaches Toward Their Reduction, Washington, DC, 23-24 Sep.1987;Sponsored by AGARD.
HARFORD J. J.; LAWRENCE B.
American Inst.of Architects Foundation, Washington, DC.
National Aeronautics and Space Administration, Washington, DC.
078190000; AR541605
Conference
ENG
US
NP. 5; DP. 1988.
S2713
NTIS Prices: PC A02/MF A01
An overview of the barriers to the international transfer of information, particularly in the aerospace and defense area is discussed.The role of the professional society, motives, and types of barriers are also discussed.
70 05; 84 00; 74 00; 88 00
Communication*;Information dissemination*;Information transfer*;
International cooperation*;Problem solving*;Aerospace industry;
Economics;Information retrieval;Machine translation;Organizations,
Politics;Standards
Barriers*;NTISNASA

TIB/B89-80975/XAD
GPSG and German word order.
HAUENSCHILD C.
Technische Univ.Berlin (Germany, F.R) Projektgrkuppe Kuenstliche Intelligenz und Textverstehen.
Bundesministerium fuer Forschung und Technologie, Bonn (Germany, F.R.).
030172001
Report
ENG
DE
KIT-52
NP. 27; DP. Jun 87.
U8915
NTIS Prices: PC E07
BMFT 10 13207-1
In this paper, the main concern is raising questions rather than giving answers.The starting point is Hans Uszkoreit's revised version of the LP (linear precedence) component within the formalism.The author discusses some problems of Uszkoreit's approach that result from the fact that the whole complex phenomenon of German word order is described at a unique level of linguistic representation.He then proposes a somewhat speculative solution to some of these problems, which is based on a multi-level approach to analysis and generation within the context of **machine** **translation** (which is the setting of the project KIT/NASEV and its successor KIT/FAST).(orig.).(Copyright (c) 1989 by FIZ.Citation no.89:080975.).
92 04
Phrase structured grammars*;German word order*;Machine translation*;Linear precedence component
Foreign technology*;NTISTFFIZ:NTISFNGE

N89-19919/4/XAD
From ALGOL 60 to Ada: Problems, Solutions, Feasibility.
HUIJSMAN R. D.; VANKATWIJK J.; PRONK C.; TOETENEL W. J.
Technische Hogeschool Delft (Netherlands).Dept.of Mathematics and Informatics Computer Science.
National Aeronautics and Space Administration, Washington, DC.
016196068; TJ479965
Report
ENG
NL
REPT-88-41
NP. 47; DP. 1988.
S2712
NTIS Prices: PC A03/MF A01
Mechanical conversion of Algol 60 programs into Ada programs was studied.Major problem areas include handling goto's, handling procedures and parameters, and interaction with the environment.A large number of Algol 60 constructs turn out to be hard or even impossible to map.Tests, mainly in scientific computation, suggest that, depending on the amount of effort to be put into the project, between 80 pct and 90 pct of the source code can be translated mechanically.Translation based only on lexical and syntactical information is possible for about 50 pct of the source text.Taking semantics into account augments the percentage to 80 pct to 90 pct The remaining 10 pct to 20 pct of the code can be translated partly if certain forms of idiom are taken into account, the rest of the code being not mechanically translatable.The inability to translate 20 pct of the source is even worse than it seems.A manual translation of the remaining code often requires a complete restructuring of the program, including those parts that could be translated mechanically.Since translation is manual to a considerable extent, maintenance is also problematic and only possible when applied to the resulting Ada programs.Unfortunately, the readability and recognizability of the latter are seriously impaired by the consequences of partial manual translation.
62 02
Ada Programming language*;Algol*;Machine translation*;Computer systems performance;Cost effectiveness;Software tools
Foreign technology*;NTISNASAE;NTISFNNL

DE88756558/XAD
Investigation of Knowledge Structure of Nuclear Data Evaluation Code.
UENAKA J.; KAMBAYASHI S.
Japan Atomic Energy Research Inst., Tokyo.
014802000; 3413000
Report
JAP
JP.
JAERI-M-88-143
In Japanese; U.S. Sales Only; NP. 71; DP. Aug 88.
U8915
NTIS Prices: PC A04/MF A01

In this report, investigation results of knowledge structure in a nuclear data evaluation code are described.This investigation is related to the natural language processing and the knowledge base in the research theme of Human Acts Simulation Program (HASP) begun at the Computing Center of JAERI in 1987.By using a **machine** **translation** system, an attempt has been made to extract a deep knowledge from Japanese sentences which are equivalent to a FORTRAN program CASTHY for nuclear data evaluation.With the knowledge extraction method used by the authors, the verification of knowledge is more difficult than that of the prototyping method in an ordinary AI technique.In the early stage of building up a knowledge base system, it seems effective to extract and examine knowledge fragments of limited objects.(ERA citation 14:014093).
77 00; 46 00; 62 00
Nuclear Data Collections*;Artificial Intelligence;C Codes; Evaluation;Human Factors;Programming;Programming Languages; Simulation;Translators
Foreign technology*;Knowledge bases Artificial intelligence *;ERDA 990210;ERDA 990300;NTISDEE;NTISFNJA;NTISLNJAP

AD-A203 445/2/XAD
Comparative Study of Japanese and English Sublanguage Patterns.
TELLER V., KOSAKA M.; GRISHMAN R.
New York Univ., NY.
008702000; 257250
Report
ENG
US
PROTEUS-M-15
Memorandum rept.; NP. 13; DP. Jun 88.
U8911
NTIS Prices: PC A03/MF A01
N00014-85-K-0163
As part of a project to develop a Japanese-English **machine** **translation** system for technical texts within a limited domain, we conducted a study to investigate the roles that sublanguage techniques and operator-argument grammar would play in the analysis and transfer stages of the system.The data consisted of fifty sentences from the Japanese and English versions of the FOCUS Query Language Primer, which were decomposed into elementary sentence patterns.A total of 187 pattern instances were found for Japanese and 191 for English.When the elements of these elementary sentences were classified and compared with their counterparts in the other language, we identified 43 word classes in Japanese and 43 corresponding English word classes.These word classes formed 32 sublanguage patterns in each language, 29 of which corresponded to patterns in the other language.This paper examines in detail these correspondences as well as the mismatches between sublanguage patterns in Japanese and English.The high level of agreement found between sublanguage categories and patterns in Japanese and English suggests that these categories and patterns can facilitate analysis and transfer.Use of operator-argument grammar, which incorporates operator trees as an intermediate representation, substantially reduces the amount of structural transfer needed in the system.(EDC).
92 04
Linguistics*;English language*;Japanese language*;Machine translation*;Operators Mathematics;Comparison;Grammars;Patterns; Matching;Structural properties;Technical writing;Information transfer;Trees;Words Language
Sublanguage patterns;Sentence patterns;Operator argument grammars; NTISDODXA

AD-A203 444/5/XAD
Domain Modeling for Language Analysis.
GRISHMAN R.
New York Univ., NY.
008702000; 257230
Report
ENG
US
PROTEUS-M-14
Memorandum rept.; Sponsored in part by Grant NSF-DCR-85-01843; NP. 9; DP. Feb 88.
U8911
NTIS Prices: PC A02/MF A01
N00014-85-K-0163
In section 2 of this paper we briefly characterize our notion of understanding a text.In section 3 we give an overview of the system we have constructed for analyzing equipment failure messages, and indicate the points at which it makes use of domain information.We then turn in section 4 to the domain model itself, and describe how it provides the information needed by language analysis.We close with brief sections relating our work to other work on discourse analysis and discussing how our system's coverage may be broadened.(FR).
92 04
Language translation*;Models;Data storage systems;Information sciences
Machine translation;NTISDODXA;NTISNSFG

AD-A203 096/3/XAD
Research in Natural Language Processing January 15, 1985 - September 15, 1987.
GRISHMAN R.
New York Univ., NY.
008702000; 257250
Report
ENG
US
PROTEUS-M-12
Memorandum rept., Sponsored in part by Grant NSF-DCR-85-01843, NP. 5; DP. Nov 87.
U8910
NTIS Prices: PC A02/MF A01
N00014-85-K-0163; N00014-85-K-2028
This report describes research done by the PROTEUS Project at New York University during the period January 15, 1985 to September 15, 1987.All of the activities described below were supported in part by the Strategic Computing Program of the Defense Advanced Research Projects Agency under Contract N00014-K-0163 from the Office of Naval Research.The PROTEUS Syntactic Analyzer is intended to provide an efficient, easy-to-use base for the various experiments in computational linguistics.The basic, long-term objective, as part of the Strategic Computing Program in Natural Language Processing, is to develop the technology necessary for the robust automated processing of messages containing natural language narrative.One aspect of the development of such language processing systems is the incorporation of detailed domain knowledge and the effective use of such knowledge in language analysis.The research has focused on one type of message, CASREPs (equipment casualty reports), on developing detailed domain knowledge (a model of the equipment), and on using this knowledge for language understanding.Keywords: Text processing, Parallel parsing, Semantics.(kr).
62 00
Computational linguistics*;Message processing*;Natural language*; Text processing*;Analyzers;Automation;Information processing; Language translation;Models;Parallel orientation;Parsers;Proteus; Reports;Semantics;Syntax
PROTEUS Project:NTISDODXA;NTISNSFG

PB89-150460/XAD
Integrated **Machine** **Translation** System PIVOT,
MURAKI K.; ICHIYAMA S.; OKAZAKI Y.; NAGAO Y.; AOKI Y.
Nippon Electric Co.Ltd., Tokyo.
020576000
Report
JAP
JP
Text in Japense. Included in NEC Technical Jnl., v41 n12 p70-76 1988; NP. 7; DP. 1988.
U8908
NTIS Prices: (Order as PB89-150452I, PC E07/MF A01)
The Integrated **Machine** **Translation** System PIVOT is a **machine** **translation** system using a knowledge base that accumulates knowledge of what is to be translated.The use of the epochal PIVOT system for **machine** **translation** system (inter mediate expression by conceptual structure) permits high quality translation and realizes integrated **machine** **translation** from Japanese to English and vice versa.Expansion to **machine** **translation** of multiple languages will be easy The paper describes the characteristics and functionality of the Integrated **Machine** **Translation** System PIVOT which was developed for reward such object.
92 04
Machine translation*
PIVOT system*;Knowledge based systems;NTISTFNEC;NTISFNJA;NTISLNJAP

PB89-150452/XAD
NEC (Nippon Electric Company) Technical Journal, Vol 41, No.12, October 1988.Special Issue: On Application Software.
Nippon Electric Co.Ltd., Tokyo.
020576000
Report
JAP
JP
Text in Japanese with English abstracts. See also PB89-150460, PB89-150478 and Volume 41, Number 11, PB89-150445.Portions of this document are not fully legible. Color illustrations reproduced in black and white; NP. 196; DP. 1988.
U8908
NTIS Prices: PC E07/MF A01
Special issue on application software. Office application system;Software development support/operation management systems;AI application systems;Scientific and engineering application systems;CAE systems.
62 02; 92 04; 70 03; 70 02; 41 02, 94 07, 99 00, 89 02; 63 07
Application programs Computers*;Machine translation*;Management information systems*;Decision making;Information retrieval; Accounting;Records management,Operations,Computer graphics, Identification systems,Engineering;Operations research,Chemistry, Architecture
Foreign technology*,Office automation*;Expert systems*,Computer aided engineering*,Science fields*,Computer applications,NTISTFNEC, NTISFNJA;NTISLNJAP

PB89-128276/XAD
Current Status of Japanese to English **Machine** **Translation**.Repor. to Congress,
KUSUDA T.
Department of Commerce, Washington, DC.Office of Japanese Technical Literature.
001948069
Report
ENG
US
NP. 18P; DP. Jul 88.
U8904
NTIS Prices: PC E03/MF A01
The primary obstacle to access to the Japanese technical literature is the Japanese language.Manual translation of Japanese technical material tends to be very expensive and, especially in specialized technical fields, is often inaccurate.Machine-aided translation (MT) offers the hope of eventually gaining a much

broader access to Japanese scientific and technical literature.The report to the U.S.Congress assesses the present state of Japanese-to-English MT.Consideration is given to the MT process itself, and to current activities in the U.S., Japan and Europe.Attention is also given to the status of optical Japanese character recognition devices as an input method for MT systems.
92 04; 62 06; 70 05
Machine translation*;Japanese language*;English language*;Input;
Optical character recognition device;Technology transfer
NTISTFNTI

AD-A199 183/5/XAD
Principle-Based Parsing for **Machine** **Translation**.
DORR B. J.
Massachusetts Inst.of Tech., Cambridge.Artificial Intelligence Lab.
001450241; 407483
Report
ENG
US
AI-M-947
Memorandum rept.: Sponsored in part by Grant NSF-DCR85-552543; NP. 18; DP. Dec 87.
U8903
NTIS Prices: PC A03/MF A01
N00014-80-C-0505; N00014-85-K-0124
Many syntactic parsing strategies for **machine** * translation** systems are based entirely on context-free grammars.These parsers require an overwhelming number of rules;thus, translation systems using rule-based parsers either have limited linguistic coverage, or they have poor performance due to formidable grammar size.This report shows how a principle-based parser with a co-routine design improves parsing for translation.The parser consists of a skeletal structure-building mechanism that operates in conjunction with a linguistically based constraint module, passing control back and forth until a set of underspecified skeletal phrase-structures is converted into a fully instantiated parse tree.The modularity of the parsing design accommodates linguistic generalization, reduces the grammar size, allows extension to other languages, and is compatible with studies of human language processing.Keywords: Natural language processing, Interlingual translation, Parsing, Subroutines, Principles vs.Rules, Co-routine design, Linguistic constraints.(edc).
92 04; 62 02
Machine translation*;Parsers*;Syntax*;Control;Grammars;Humans;
Language;Linguistics;Modular construction,Natural language;
Processing;Strategy;Subroutines
Coroutine design;Principles based parsers;NTISDODXA;NTISNSFG

PB88-253554/XAD
MELTRAN: A Japanese-English **Machine**-**Translation** System Using the MELCOM PSI II,
DASAI T., SUZUKI K., KIYOHARA R , MARUYAMA F., YOSHITAKE J.
Mitsubishi Electric Corp., Tokyo (Japan).
076350000
Report
JAP
JP
Text in Japanese; Included in Mitsubishi Denki Giho, v62 n5 p57-60 1988; NP. 5; DP. 1988.
U8824
NTIS Prices: (Order as PB88-253521, PC E05/MF A01)
The MELTRAN system achieves high-quality translation through interpretation rules that identify special usages as well as general linguistic constructions.New rules can be added to customize the system for specific applications.The system translates 10,000 words/h.(Copyright (c) 1988, Mitsubishi Electric Corporation.).
92 04
Machine translation*,Japanese language*;English language*;
Automatic language processing;Linguistics
MELCOM PSI II computers;NTISDFMIT;NTISFNJA;NTISLNJAP

AD-A197 356/9/XAD
Lexical Conceptual Approach to Generation for **Machine** **Translation**.
DORR B. J.
Massachusetts Inst.of Tech., Cambridge.Artificial Intelligence Lab.
001450241; 407483
Report
ENG
US
AI-M-1015
Memorandum rept. Sep 87-Jan 88; NP 26, DP. Jan 88.
U8824
NTIS Prices: PC A03/MF A01
N00014-85-K-0124
Current approaches to generation for **machine** **translation** make use of direct-replacement templates, large grammars, and knowledge-based inferencing techniques.Not only are rules language-specific, but they are too simplistic to handle sentences that exhibit more complex phenomena.Furthermore, these systems are not easily extendable to other languages because the rules that map the internal representation to the surface form are entirely dependent on both the domain of the system and the language being generated.Finally an adequate interlingual representation has not yet been discovered;thus, knowledge-based inferencing is necessary and syntactic cross-linguistic generalization cannot be exploited.This report introduces a plan for the development of a theoretically based computational scheme of natural language generation for a translation system.The emphasis of the project is the mapping from the lexical conceptual structure of sentences to an underlying or base syntactic structure called deep structure.This approach tackles the problems of thematic and structural divergence, i.e., it allows generation of target language sentences that are not thematically or structurally equivalent to their conceptually equivalent source language counterparts.Two other more secondary tasks, construction of a dictionary and mapping from deep structure to surface structure, will also be discussed.The generator operates on a constrained grammatical theory rather than on a set of surface level tranformations.(kr).
92 04; 62 03
Systems approach*;Lexicography*;Machine translation*;Natural language*;Computations;Dictionaries;Grammars;Internal;Surfaces;
Syntax;Theory;Translations;Words Language
NTISDODXA

PB88-225446/XAD
NSF (National Science Foundation) Tokyo Reports.Report Memoranda No.90-102.
National Science Foundation, Tokyo (Japan).Tokyo Office.
National Technical Information Service, Springfield, VA. Office of International Affairs.
091981001
Report
ENG
JP
See also PB88-225438 and PB88-225453. Sponsored by National Technical Information Service, Springfield, VA. Office of International Affairs, NP 304; DP. Jun 88.
U8821
NTIS Prices: PC E11/MF A01
The document presents Report Memoranda issued by the Tokyo Office of the U.S.National Science Foundation (NFS) during the first half of 1986.The Memoranda included in the volume are: 1985 Survey of Research and Development in Japan;Japan Key Technology Center;Directory of Japanese Company Laboratories Willing to Receive American Researchers;Japanese S&T Budget for Japanese Fiscal Year 1986;Japanese **Machine** **Translation** Efforts -- A Look at Three Selected MT Systems;A Visit with Dr.Jiro Kondo, President, Sicence Council of Japan;Proposed New Law to Encourage Industry/Government Cooperation in Science and Technology;STA and RIKEN to Launch 'International Frontier Research System';Japan's Key Technology Center Selects Twenty-five R&D Projects for Capital Investment;Recommendations of Monbusho's Science Council for Promotion of Bioscientific Research;A Visit with Dr.Saburo Nagakura, Director-General, Institute of Molecular Science, Okazaki National Research Institutes;Monbusho's R&D Budget for JFY 1986;Japan's Science and Technology Agency R&D Budget for JFY 1986.
70 05
Japan*;Research projects;Laboratories;Budgeting
Foreign technology*;NTISTFNTI;NTISFNJA

PB88-211073/XAD
Directory of Japanese Technical Resources in the United States, 1988.
National Technical Information Service, Springfield, VA.Office of International Affairs.
055665003
Report
ENG
US
ISBN-0-934213-13-5
Supersedes PB87-205258; NP. 165P; DP. 1988.
U6821
NTIS Prices: PC$36.00/MF A01
Developed in response to the Japanese Technical Literature Act of 1986, the Directory has been divided into four parts.The first part contains an alphabetical list of commercial services that collect, abstract, translate or disseminate Japanese technical information.Following this are two indices, one by area of specialization and one by state.The second part lists Government agencies with programs and services involving Japanese technical information.The third contains libraries in both the public and private sectors that have extensive holdings of Japanese technical information The final part cites Japanese technical documents translated at Federal expense which are available to the public.In addition to these directories, the publication also includes background articles: (1) universities that have initiated programs to provide undergraduate and graduate students, as well as experienced scientists and engineers, with sufficient proficiency in Japanese to enable them to take advantage of the large amount of untranslated material emanating from Japan;(2) the status of Japanese-to-English **machine** **translation** projects in the United States, Europe and Japan;(3) U.S.Government efforts to implement the Japanese Technical Literature Act;(4) follow-up on two case studies reported in the 1987 Directory;(5) a private sector view of America's readiness to take advantage of Japanese technology.
88 0[illegible] 88 01; 70 05
Directories*;Science*;Engineering*;Japan*;Technical reports;
Libraries;Translations
Technical information centers*;Science and technology*;Technology transfer*,Information sources*;Technology utilization;Government agencies;NTISTFNTI

AD-A195 281/1/XAD
UNITRAN (UNIversal TRANslator): A Principle-Based Approach to **Machine** **Translation**.
DORR B. J.
Massachusetts Inst.of Tech . Cambridge.Artificial Intelligence Lab.
001450241; 407483
Report
ENG
US
ARI-TR-1000

Technical rept. Sep 84-May 87; Sponsored in part by Grant
NSF-DCR85-552543; NP. 314; DP. Dec 87.
U8821
NTIS Prices: PC A14/MF A01
N00014-80-C-0505; N00014-85-K-0124

This report presents an approach to natural language translation that relies on principle based descriptions of grammar rather than rule-oriented descriptions.The model that has been constructed is based on abstract principles as developed by Chomsky (1981) and several other researchers working within the 'Government and Binding' (GB) framework.The approach taken is 'interlingual', i.e., the model is based on universal principles that hold across all languages;the distinctions among languages are then handled by settings of parameters associated with the universal principles.The design of the UNITRAN (UNIversal TRANslator) system is such that a language may be described by the same set of parameters that specify the language in linguistic theory.Because of the modular nature of the model, the interaction effects of universal principles are easily handled by the system;thus, the programmer does not need to specifically spell out the details of rule applications.Because only a small set of principles covers all languages, the unmanageable grammar size of alternative approaches is no longer a problem.Keywords: Natural language processing, Interlingual **machine** **translation**, Co-routine design, Principles and parameters, Parsing, Thematic substitution.

92 04
Machine translation*;Natural language*;Abstracts;Grammars;
Interactions;Language;Language translation;Linguistics;Parameters,
Parsers;Processing;Sizes Dimensions;Substitutes;Theory
Natural language processing;Interligual machine translation,
Thematic substitution;NTISDODXA;NTISNSFG

AD-A194 653/2/XAD
Barriers to Information Transfer and Approaches Toward Their Reduction, Conference Proceedings of the Technical Information Panel Specialists' Meeting Held in Washington, DC on 23-24 September 1987.
Advisory Group for Aerospace Research and Development, Neuilly-sur-Seine (France).
056102000; 400043
Report
ENG
ZZ
AGARD-CP-430
Conference proceedings; For sales information of individual items see AD-P005 721 thru AD-P005 730; NP. 108; DP. Mar 88.
U8820
NTIS Prices: PC A06/MF A01

Contents: Barriers to the International Transfer of Information in Aerospace and Defense;Linguistic and Cultural Barriers to the Transfer of Information;Political and Economic Barriers to Information Transfer;Linguistic and Technical Aspects of **Machine** **Translation**;Information Retrieval Systems Evolve - Advances for E' ier and More Successful Use;Information Technology to Facilitate oup Interaction;Words: Key or barriers to Information Transfer;Linguistic Barriers:Translation Problems;Technical Change Needs Organizational Change,and Using Standards to Break Down Information Transfer Barriers.

88 02; 70 05
Information transfer*;Aerospace industry*;Research management*,
Symposia*;Barriers
Foreign technology*;Compilation Reports;NATO furnished;NTISDODXA;
'NTISFNZZ

PB88-230073/XAD
Machine **Translation** of On-Line Searches in Japanese Databases,
SIGURDSON J.; GREATREX R.
Lund Univ.(Sweden).Research Policy Inst.
016503007
Report
ENG
SE
ISBN-91-86002-62-7
NP. 133; DP. 1987.
U8819
NTIS Prices: PC E07/MF E07

Contents: The information industry and databases in Japan (Availability of information on Japan in the West, The development of the Japanese information industry, The use of databases in Japan, The development of the database industry in Japan);**Machine** **translation** of Japanese-an introduction (Research and development on **machine** **translation** in Japan, Source tests, The Japan-info project in Europe, **Machine** **translation** systems in Japan, Assessment of **machine** **translation** in connection with searches in Japanese-language databases);Description of selected major databases in Japan (Description of databases vendours, Nikkei T- acom);**Machine** **translation** systems (The basic linguistic problems, Levels of translation, **Machine** **translation** in practice, The combination on-line search/**machine** **translation**, Japanese vs.Western usage, Human translation vs.**machine** **translation**, Commercially available **machine** **translation** systems, Experiences of MT system users: Honda R&D Center, Experiences of MT system users: Daiwa Securities Group);Examples of **machine** **translation** (The automatic translation of database searches, Examples of translations achieved with pre-editing, Comparative translations)

62 02; 95 06
Machine translation*,English language*,Japanese language*,Grammars,
Computational linguistics,Data processing,Models,Prototypes.
Natural language;Data bases
Foreign technology*;NTISTFRPI;NTISFNSW

AD-A193 631/9/XAD
UNITRAN: An Interlingual **Machine** **Translation** System.
DORR B. J.
Massachusetts Inst.of Tech., Cambridge.Artificial Intelligence Lab.
001450241; 407483
Report
ENG
US
AI-M-999
Memorandum rept.; NP. 15; DP. Dec 87.
U8819
NTIS Prices: PC A03/MF A01
N00014-80-C-0505; N00014-85-K-0124

Machine'**translation** has been a particularly difficult problem in the area of Natural Language Processing for over two decades.Early approaches to translation failed, partly because interaction effects of complex phenomena made translation appear to be unmanageable.Later approaches to the problem have been more successful but are based on many language-specific rules of a context-free nature.To try to capture all of the phenomena allowed in natural languages, context-free rule-based systems require an overwhelming number of rules;thus, such translation systems either have limited linguistic coverage, or they have poor performance due to formidable grammar size.This report presents an implementation of an alternative approach to natural language translation.The UNITRAN (Universal Translator) system relies on principle-based descriptions of grammar rather than rule-oriented descriptions.The approach taken is interlingual, i.e., the model is based on universal principles that hold across all languages;the distinctions among languages are then handled by settings of parameters associated with the universal principles.The grammar is viewed as a modular system of principles rather than a large set of ad hoc language-specific rules.Interaction effects of linguistic principles are handled by the system so that the programmer does not need to specifically spell out the details of rule applications.Only a small set of principles covers all languages;thus, the unmanageable grammar size of alternative approaches is no longer a problem.

92 04
Machine translation*;Grammars;Interactions,Language;Language
translation;Linguistics;Modular construction;Natural language;
Processing;Sizes Dimensions;Translations
Foreign technology*;UNITRAN translation system*;NTISDODXA

AD-A192 449/7/XAD
Automatic Error Recovery in a Fast Parser,
GRAY R. W.
Colorado Univ.at Boulder.Dept.of Computer Science.
068646006; 409496
Report
ENG
US
NP. 10; DP. 1987.
U8817
NTIS Prices: PC A02/MF A01

Although parser generators have provided significant power for language recognition tasks, many of them are deficient in error recovery.Of the ones that do provide error recovery, many of these produce unacceptably slow parsers.I have designed and implemented a parser generator that produces fast, error recovering parsers.For any input, the error recovery technique guarantees that a syntactically correct parse tree will be delivered after parsing has completed.This improves robustness because the remaining compilation phases, such as semantic analysis, will not have to deal with infinitely many special cases of incorrect parse trees.The high speed of the parser is a result of making the code directly executable and paying careful attention to implementation details.Measurements show that the generated parser runs faster than any other parser examined, including handwritten recursive descent parsers.The cost of this fast parser with error recovery is a slight increase in space.Although this particular generator requires LL grammars, the ideas can be applied to generators taking LALR grammars.Furthermore, we give the transformations that allow one to transform may LALR grammars into equivalent LL grammars.

62 02
Parsers*;Machine translation*;Automatic;Costs;Errors;Generators;
Grammars;Guarantees;Language;Recognition;Recovery;Semantics;
Computer applications
Logic trees;NTISDODXA

DE88701242/XAD
Computerized Output of Phonetic Codes in Devanagari Script by DOT-Matrix Printers.
SOMASUNDARAM S.; SURI M. M. K.; KHATUA R.
Bhabha Atomic Research Centre, Bombay (India)
004104000; 0807000
Report
ENG
IN
BARC-1358
U.S. Sales Only; NP. 6; DP. 1987.
U8816
NTIS Prices: PC A02/MF A01

This report describes the development of a computer software for converting hex-octal, alpha-numeric and pure-alpha mode input in English into 'phonetic Devanagari characters', which can be printed through dot-matrix printers in 2 passes of print-head, along with English text in the same lines.If multilingual terminals presently available in India, are used, it requires 4 passes of print-head for printing phonetic Devanagari characters, and English text also is converted into phonetic Devanagari script during printing.Thus, the software reported in this, is an improvement over the facilities currently available in the Indian

market.9 tables, 2 refs.(Atomindex citation 19:019387).
92 04
Identification Systems*;Personnel*;Computer Codes;India;Radiation Doses
Foreign technology*;Multilingualism*;Computer software*;Machine translation*;ERDA 990220;English language;Devanagari;Printers;Output;NTISINIS;NTISFNIN

TIB/A88-80799/XAD
Analysis of verb addition in PHRASEG.
Verbzusatzanalyse in PHRASEG.
SCHMITZ K. D.
Universitaet des Saarlandes, Saarbruecken (Germany, F.R.).Sonderforschungsbereich 100 Elektronische Sprachforschung.
019985001
Report
GER
DE
In German,Universitaet des Saarlandes, Sonderforschungsbereich 100 - Elektrotechnische Sprachforschung, Teilprojekt K. Dokumentation, no. 5.3; NP. 19; DP. 26 Mar 85.
U8815
NTIS Prices: PC E07
The present state of verb addition analysis is described in the context of segmenting natural language sentences within the translation system SUSY.Exact data are given on treating the verb addition in SUSY (the operator carries out the sentence segmenting so that the verb and the verb addition are combined), on verb addition analysis in rules and tables and for expanding the program structure (program structure plan, unchanged or new sub-programs).The practice of verb addition analysis (operator PHRASEG) is made clear from some alphanumeric examples (different sentences and their segments, word order, nodes, main sentences).(HWJ).(TIB: RO 3907 (5.3).) (Copyright (c) 1988 by FIZ.Citation no.88.080799.).
92 04
Sentence structure*;Linguistics*;Translation;Natural language;Sentences,Rules;Segments,Computer programs,Combinatorial analysis
Foreign technology*,SUSY system;Natural language;Machine translation;NTISTFFIZ;NTISFNGE;NTISLNGER

TIB/A88-80027/XAD
The SUSY dictionaries as a linguistic basis of knowledge.
SUSY-Lexika als linguistische Wissensbasis.
LUCKHARDT H. D.
Universitaet des Saarlandes, Saarbruecken (Germany, F.R.).Sonderforschungsbereich 100 Elektronische Sprachforschung.
019985001
Report
GER
DE
In German,CL-Report Berichte ueber Arbeiten im Bereich der Computerlinguistik aus dem SFB 100, no. 7. NP. 49; DP. Aug 85.
U8810
NTIS Prices: PC E07
Knowledge bases in the form of semantic networks and of (word) expert systems are available for organising knowledge.However, they are limited thematically.If one does not want to limit the operating area of machine parsers, one must fulfil certain requirements.The data must not be isolated from each other: The dictionary and structural connection between the source and target languages of the system must be produced;the syntactial and semantic information must be sufficient for analysis and translation;and the data organisation must be easy to expand and to correct.There was a detailed examination of the extent to which the SUSY dictionary system fulfils these requirements.Data on this concern the organisation of statistical linguistic knowledge and the linguistic knowledge of the SUSY dictionaries (analysis, semantic, transfer and synthesis dictionaries).(HWJ).(TIB: RO 2852 (7).) (Copyright (c) 1988 by FIZ.Citation no.88:080027.).
92 04
Computational linguistics*;Machine translation*;Semantics,Parsing algorithms;Dictionaries;Encyclopedias;Expert systems,Logic programming,Syntax
Foreign technology*;Knowledge based sytems*;Expert systems*;NTISTFFIZ;NTISFNGE;NTISLNGER

PB88-100920/XAD
Overview of the Conceptual Language Koto,
WEIGAND H.
Vrije Univ., Amsterdam (Netherlands).Subfaculteit Wiskunde en Informatica.
019507011
Report
ENG
NL
IR-112
NP. 29; DP. Oct 86.
U8803
NTIS Prices: PC E03/MF A01
The paper presents a preliminary overview of the conceptual language Koto, developed for knowledge base applications.Its underlying assumption is that knowledge representation is natural language bound.Koto is a way to represent different levels of information present in a sentence.Inference rules are presented that allow for syllogistic reasoning.Koto has several applications: (1) conceptual modelling;(2) knowledge representation;and, (3) **machine** **translation**.
62 02; 95 06; 92 04
Programming language*;Machine translation,Natural language
Foreign technology*;KOTO programming language*;Knowledge representation*;NTISTFNPO;NTISFNNL

PB88-112065/XAD
Electronic Databases in Japan: An Information Resource to be Reached On-Line.
SIGURDSON J.; GREATREX R.
Lund Univ.(Sweden).Research Policy Inst.
016503007
Report
ENG
SE
ISBN-91-86002-59-7
NP. 138P; DP. 1986.
U8802
NTIS Prices: PC E07/MF E07
It is shown in the report that the database industry in Japan is on the verge of becoming a more dynamic sector fueled by the demand for more selective and timely information but also by new developments including those of **machine** **translation**.The establishment of a new promotion organization and initiatives by MITI including new financial schemes by the Japan Development Bank give further indication that the Japanese database industry will undergo considerable growth.Against this background the Research Policy Institute has carried out exploratory searches in a few of the major databases in Japan covering industrial, technical, economic and political fields.The report clearly demonstrates that on-line searches in Japanese language databases can with relative ease be carried out where good telephone lines are available.It has also been demonstrated that such searches make it possible to obtain critical information more selectively and more quickly than traditional ways of obtaining the same information.It is also demonstrated that Japanese databases contain pertinent information which may ordinarily be difficult to obtain.Table of Contents: The database industry in Japan;Searching in Japanese databases,Comments on some databases for economics, and science and technology;Actual searches in selected databases;Online database searching procedure,Evaluation of searches;**Machine** **translation**;Major Japanese databases - Selective listing;Cost of on-line searching;List of manuals and thesauruses.
88 04; 70 05; 92 04
Information systems*;Information retrieval*;Searching*;Machine translation*;Japan*,Japanese language;Translating;Economics;Political systems;Finance
Foreign technology*;Data bases*;Access*;Data base industry*,Science and Technology;Information sources;On line systems;Industrial sector;NTISTFRPI;NTISFNSW

PB87-229282/XAD
NEC (Nippon Electric Company) Journal, Vol.39, No.8, 1986.,
Nippon Electric Co.Ltd., Tokyo.
020576000
Report
JAP
JP
Text in Japanese with English abstracts and illustration notes. See also PB87-229260.Portions of this document are not fully legible; NP. 95; DP. 1986.
U8723
NTIS Prices: PC E04/MF A01
Remarks for special issue on artificial intelligence;NEC's approach to artificial intelligence and related product overview;AI oriented computer.The programming environment of AI languages;A tool for expert systems;An expert system of computer operation and utilization,Integrated **machine** **translation** system PIVOT;Fundamental research on artificial intelligence;Researches on artificial intelligence applications;Problems in artificial intelligence.
62 02
Artificial intelligence*;Computers;Computer programming
Foreign technology*;Expert systems;Translators;NTISTFNEC,NTISFNJ,NTISLNJAP

PB87-866604/XAD
Computerized Multilingual Translation and Manipulation.January 1983-September 1987 (Citations from The Computer Database).
National Technical Information Service, Springfield, VA.
055665000
Report
ENG
US
Rept. for Jan 83-Sep 87; NP. 48P; DP. Sep 87.
U8721
NTIS Prices: PC N01/MF N01
This bibliography contains citations concerning machine assisted language translation and computers that manipulate several languages with dissimilar alphabets.Software packages that translate Chinese, French, German, Italian, Japanese, Spanish, and Arabic to English are discussed, along with software that translates English to other languages.Word processors and computers that manipulate Hebrew/English, Arabic/English, Arabic/French, and English/Chinese characters are included.(Contains 97 citations fully indexed and including a title list.).
92 04; 88 05
Bibliographies*;Machine translation*;Automatic language processing*,Computational linguistics*;Syntax;Semantics;Artificial intelligence,Translating,English language,Russian language,German language;Chinese language;Japanese language
Foreign languages*,Natural language,Vocabulary;NTISPSCOM;NTISNERACD

PB87-866349/XAD
Machine **Translation**: Foreign Language Translation and Natural Language Understanding.January 1970-August 1987 (Citations from the NTIS Database).
National Technical Information Service, Springfield, VA.

055665000
Report
ENG
US
Rept. for Jan 70-Aug 87; Supersedes PB81-806507; NP. 41P; DP. Sep 87.
U8721
NTIS Prices: PC N01/MF N01
This bibliography contains citations concerning research and development of machine/mechanical foreign language translation by computer.Topics include syntactic and semantic translations, natural language representation and understanding, knowledge based systems, language manuals for ideographic machines, systran **machine** **translation**, mathematical linguistics and logic, foreign technologies and language translation, processes for question answering, and Chinese lexicography and romanization.Methods and systems for translations of Russian, German, Chinese, and Japanese to English are presented.(This updated bibliography contains 96 citations, 25 of which are new entries to the previous edition.).
92 04; 62 00; 88 05
Bibliographies*;Machine translation*;Automatic language processing*;Computational linguistics*;Syntax;Semantics;Artificial intelligence;Translating;English language;Russian language;German language;Chinese language;Japanese language
Foreign languages*;Natural language;Vocabulary;NTISNTISN;NTISNERACD

ED-281 519
Arabization of a Full-Text Database Interface.
FAYEN E. G.
Georgia Inst.of Tech., Atlanta.
Agency for International Development (IDCA), Washington, DC.
010263000
Report
ENG
US
Available from ERIC Document Reproduction Service (Computer Microfilm International Corporation), 3900 Wheeler Ave., Alexandria, VA 22304-5110; NP. 5; DP. 1986.
U8722
NTIS Prices: Not available NTIS
NEB-0016-00-2092
The 1981 design specifications for the Egyptian National Scientific and Technical Information Network (ENSTINET) stipulated that major end-user facilities of the system should be bilingual in English and Arabic.Many characteristics of the Arabic alphabet and language impact computer applications, and there exists no universally accepted character encoding scheme equivalent to the ASCII standard for Latin alphabets.In order to overcome the native language barrier in the system, a native larguage interface to existing software was developed.The Arabic language software functions include an Arabic editor running under the UNIX operating system, an Arabic database search facility, and electronic mail, which were implemented for peripheral devices using the CODAR/UFD Arabic character encoding scheme.The Arabic database search facility has been developed by arabizing BRS/Mate, a menu-oriented front end to the native mode of Mini-Micro BRS/Search, a full text, state-of-the-art information management software system.Five references are provided.(MES).
88 02
Arabic*;Developing nations*;Man machine systems*;Online searching*. Online systems*;Character recognition;Computer software;Databases, Developed nations;English;Foreign countries;Machine translation; Scientific and technical information;Technology transfer;Users Information
Egyptian National Scientific Tech Info Network*;Nonroman Scripts*, Bibliographic Retrieval Services;Character Encoding;Egypt;Full Text Databases;Software Design;User Cordial Interface;NTISHEWERI

PB87-863098/XAD
Chinese and Japanese Language Translation by Computer.January 1975-June 1987 (Citations from the INSPEC Information Services for the Physics and Engineering Communities Database).
National Technical Information Service, Springfield, VA.
055665000
Report
ENG
US
Rept. for Jan 75-Jun 87; NP 34P; DP. Jul 87.
U8718
NTIS Prices: PC N01/MF N01
This bibliography contains citations concerning research and development of computer hardware and software for the language translation of Chinese and Japanese.Computer technology in character recognition, sentence analysis, text input and output systems, automatic language translation systems for personal computers, and character generation and analysis are discussed.Translation techniques for Chinese-to-Japanese, Chinese-to-English, and Japanese-to-English are presented.Applications in business, utilities management, and library automation are included (Contains 60 citations fully indexed and including a title list.)
92 04; 88 05
Bibliographies*;Machine translation*;Chinese languages;Japanese languages;Automatic language processing;Input output devices Computers;Computer systems hardware;Computer systems programs, Character recognition
Chinese language translation*,Japanese language translation*, NTISNTISH;NTISNERACD

AD-A180 138/4/XAD
Method for the Automatic Translation of Algorithms from a High-Level Language into Self-Timed Integrated Circuits.
KELEM S. H.
Aerospace Corp., El Segundo, CA.Computer Science Lab.
Space Div., Los Angeles AFB, CA.
000512051; 417388
Report
ENG
US
TR-0084A(5920-03)-1
Technical rept.; NP. 13; DP. 30 Sep 86.
U8716
NTIS Prices: PC A02/MF A01
F04701-85-C-0086
SD-TR-86-60
A method for generating custom self timed integrated circuits (ICs) from algorithmic descriptions of the desired circuits.The goal is to quickly produce prototype integrated circuit masks that implement various algorithms and data types in order to evaluate the IC power, delay, and area characteristics.A topology and behavior preserving mapping is used to perform the translation from constructs in the function language to mask primitives.Keywords: Algorithms;Integrated circuits masks;Self timed integrated circuits;High level language;Compilers;Translator;Algol 68;Templates.
49 09
Integrated circuits*;Computer aided design*;Algorithms;Compilers. High level languages;Machine translation;Prototypes;Topology, Optimization
NTISDODXA;NTISDODAF

TIB/A87-80056/XAD
Valencies and Depth Case in **Machine** **Translation**
Valenz und Tiefenkasus in der maschinellen Uebersetzung
LUCKHARDT H. D.
Universitaet des Saarlandes, Saarbruecken (Germany, F.R.).Sonderforschungsbereich 100 Elektronische Sprachforschung.
Deutsche Forschungsgemeinschaft, Bonn (Germany, F.R.).
019985001
Report
GER
DE
In German, CL-Report, no. 4; NP. 37; DP. Jul 85.
U8712
NTIS Prices: PC E07
DFG-SFB-100
In many well-known machine analysis and translation systems (MUe systems), the valencies of the verbs, adjectives and nouns are used implicitly or explicitly.The degree of explicitness is shown by the purpose for which valencies are used and what value they take up in the theoretical description.This is a report on an MUe system which has selected, implemented and tested the valency theory as the central grammatical theory.The theroetical implications of the practical use of valencies in these MUe systems are dealt with in detail.An MUe related valency theory is propounded, and the differences between obligatory and facultative factors and free data are shown.The case theory (notation variants, designed definition of case rolls) are also included in the syntactical-semantic representation.(HWJ).(TIB. RO 2852 (4)) (Copyright (c) 1987 by FIZ.Citation no.87-060056.).
92 04
Automatic language processing*;Machine translation*;Computational linguistics;Speech analysis;Valence;Syntax;Semantics
Foreign technology*;NTISFIZ;NTISFNGE;NTISLNGER

TIB/A87-80055/XAD
Information Based Generation of Speech Processing Systems.
Wissensbasierte Generierung sprachverarbeitender Systeme
THIEL M.
Universitaet des Saarlandes, Saarbruecken (Germany, F.R.).Sonderforschungsbereich 100 Elektronische Sprachforschung.
Deutsche Forschungsgemeinschaft, Bonn (Germany, F.R.).
019985001
Report
GER
DE
In German, CL-Report, no. 5; NP. 32; DP. Aug 85.
U8712
NTIS Prices: PC E07
DFG-SFB-100
In the context of a project on 'electronic speech research', the question was tackled of how natural speech systems developed on small sections of speech behave, if they are confronted by large amounts of text from wide areas of application.This problem does not occur in the development phase of many systems, as one usually works with restrung amounts of text.This is a report on solving this problem by a system, which makes access to the machine speech analysis and translation systems SUSY-II and SUSY-III easier, and which is intended to control the modification of the linguistic and strategic information contained in them.The background to the system is sketched and reference is made to theoretical bases of systems.A model developed from this is introduced (experimental generation of parts of this model).Finally, the integration of data bases in this concept is shown.(HWJ).(TIB: RO 2852 (5).) (Copyright (c) 1987 by FIZ.Citation no.87:080055.).
45 06
Automated language processing*,Computational linguistics;Machine translation,Computer systems programs;Semantics;Programming languages
Foreign technology*;Speech analysis*,NTISFIZ;NTISFNGE;NTISLNGER

AD-A175 145/2/XAD
Lexical Translator from Arabic to Latin in Pascal Environment.
ALJUHAIMAN S. S.
Naval Postgraduate School, Monterey, CA.
019895000; 251450

Thesis
ENG
US
Master's thesis; NP. 162; DP. Sep 86.
U8707
NTIS Prices: PC A08/MF A01

The Lexical translator is a program written in Turbo PASCAL to generate a Latin PASCAL source code from an Arabic PASCAL source code.The Arabic code is written under a bilingual operating system transparent to the DOS on personal computers.The bilingual operating system compatibility as well as the Arabic characters' code values is investigated.The Latin code is fed into a computer to be compiled and run with Latin interpreter (i.e., Turbo PASCAL), in an Arabic environment.(Author).

92 04; 62 02
Translations*;Alphanumeric data;Computer programs;Microcomputers; Theses
Arabic language*;Machine translation*;Languages*;Translators*; Pascal programming language;Bilingual;Computer program listings; Operating systems;Personal computers;NTISDODXA

AD-A175 122/1/XAD
Natural Language Generation: Complexities and Techniques.
MCDONALD D. D.
Massachusetts Univ., Amherst.Dept.of Computer and Information Science.
010574068; 407701
Report
ENG
US
NP. 34; DP Apr 86.
U8707
NTIS Prices: PC A03/MF A01
N00014-85-K-0017

This paper examines the nature of generation systems today, the problems they have been designed to deal with, their strengths and their weakness.Its goals to give the MT community a sense of what has been accomplished, and indirectly to show where MT researchers could consider adopting or adapting some of the AI work This work on generation need not be done by AI people alone MT can, for example, contribute to AI research on the planning-level by sharpening our collective understanding of the 'carrying capacity' of the different parts of a language through cross-language comparisons that try to fit the ideas carried by the linguistic devices of a source language into the alternative devices of a target language.At lower levels, MT as a task can provide more linguistically demanding sources for generation than most any of today's expert systems.At the same time it is clear that generation is done for very different reasons in two camps.The AI context is more like that of people dealing with each other in normal life--of which translation is not a customary part.Nevertheless, translation is a normal human capacity, and a considered comparison of the generation process in both contexts should tell us more about the nature of generation as a module within the human mind than could either by itself.

92 04
Natural language*;Machine translation*;Linguistics;Sources;Low level;Language;Targets;Artificial intelligence
Natural language generation;Counselor Project;NTISDODXA

TIB/B86-80315/XAD
KIT/NASEV or the Problems of Transfer in **Machine** **Translation**.
KIT/NASEV oder die Problematik des Transfers bei der maschinellen Uebersetzung
HAUENSCHILD C.
Technische Univ.Berlin (Germany, F R.).Projektgruppe Kuenstliche Intellizenz und Textverstehen.
030172001
Report
GER
DE
KIT-29
Text in German; NP. 16; DP. Nov 85.
U8703
NTIS Prices: PC E07

This paper deals with the problems of transfer within the framework of **machine** **translation** systems.After a brief general discussion of the role of the transfer phase in **machine** **translation**, the authors give an intuitive analysis of a typical lexical transfer problem that arises in the translation of a short German text into English.In the light of the requirements derived from that example, they propose a system of multi-level representations for source and target texts and a corresponding multi-level transfer phase for the MT project KIT/NASEV.The formalisms for the different levels of representation are illustrated on the basis of the given sample problem of lexical transfer.(Copyright (c) 1986 by FIZ.Citation no.86:80315.).

92 04; 62 00
Translators*;Machine translations*;Lexical transfer*;Transferring*; Automatic language processing*;English language;German language. Linguistics
Foreign technology*:NTISFIZ;NTISFNGE;NTISLNGER

N86-29725/6/XAD
Travail dans le Cadre d'UN Reseau de Terminologie en Matiere de Technologie de l'Espace
Research in the Framework of a Terminology Network in the Field of Space Technology.
Text in French.Presented at Infoterm Symposium (2nd).
BOUDJEDID M.
Societe Nationale Industrielle Aerospatiale, Paris (France).
National Aeronautics and Space Administration, Washington, DC
066215000; S0451674
Conference
FRE
FR
SNIAS-861-550-101; ESA-86-97178
NP. 9; DP. 1986.
S2420
NTIS Prices: PC A02/MF A01

The network implemented by a leading aerospace company is described.It is based on a terminological and linguistic coordination extended to national and international levels, a terminological data bank, and a dictionary publishing center.The integration of these activities in an industrial organization is explained.The role played by automatic translation systems is discussed.

05 02; 22 01; 88 05; 84 00; 92 04
Dictionaries*;International cooperation*;Machine translation*, Technical writing*;Terminology*;Organizing,Conferences; Documentation;Network synthesis;Words Language
Foreign technology*;Space exploration*;Data banks;NTISNASAE; NTISLNFRE;NTISFNFR

PB86-198637/XAD
THALIA-3, a Japanese-English **Machine**-**Translation** System.
OGASHI H.; SUZUKI K., OASAI T.
Mitsubishi Electric Corp., Tokyo (Japan).
076350000
Report
JAP
JP
Text in Japanese. Included in Mitsubishi Denki Giho, v60 n3 p37-40 1986; NP. 5; DP. 1986.
U8618
NTIS Prices: (Order as PB86-198587, PC E04/MF E01)

THALIA-3 is a new **machine**-**translation** system which can be operated on MELCOM-COSMO Series computers.The main purpose of its developments into English and vice versa at high speeds.It uses knowledge information technology plus semantic representation to meet this requirement It also has a basic 60,000-term dictionary, the technical vocabularly of which can be extended, and covers a wide variety of technical fields.

05 07; 09 02; 92 04
Machine translation*
Computer applications;NTISDFMIT;NTISFNJA;NTISLNJAP

PB86-198587/XAD
Mitsubishi Denki Giho, Vol.60, No. 3, 1986.
Mitsubishi Electric Corp., Tokyo (Japan).
076350000
Report
JAP
JP
Text in Japanese with English abstracts. See also PB86-198595 through PB86-198660, and PB86-198546.Color illustrations reproduced in black and white; NP. 85; DP. 1986.
U8618
NTIS Prices: PC E04/MF E01

The issue contains technical reports on: A home-use high-definition VCR;Software for the development of home electrical applicances and home-automation systems;A verification system for logic programs;THALIA-3, a Japanese-English **machine**-**translation** system Bandwidth compression of video signals by means of vector quantization;Compound semiconductor superlattice heterostructures;Recent advances in superconducting-magnet technology;The development of a three-dimensional CAD/CAM system;The SD and SAGE high-power CO2 laser excitation systems;Optical pickups for compact-disc players;Three dimensional device technology;A high-performance photomask with a molybdenum silicide film;Ergonomics in industrial design;Multibeam antennas;Magnetic heads and media for high-density disk drives;A high-resolution, high-quality thermal-printing head;A 1Mb dynamic MOS RAM.

09 02; 17 02; 14 03; 13 08; 09 05; 20 06; 62 01; 45 03; 82 03; 41 02; 49 05; 46 03
Recording instruments;Computer systems programs;Semiconductors Materials;Superconducting magnets;Lasers;Masking;Human factor engineering;Antennas;Magnetic heads;Random access computer storage
Foreign technology*;Therma printheads;NTISDFMIT,NTISFNJA;NTISLNJAP

B90000058; C90001573
International Workshop on Parsing Technologies
International Workshop on Parsing Technologies
Pittsburgh, PA, USA
28-31 Aug. 1989
Conference proceedings
Theoretical mathematical
ENG
ZZ
Carnegie Mellon Univ;Pittsburgh, PA, USA
NP. vii+467; DP. 1989

The following topics were dealt with: information based passing;disjunctive constraint satisfaction;head-driven bidirectional passing;head-driven parsing;probabilistic passing;speech recognition;dependency grammar passing;combinatory grammars;Tomita algorithm;computational complexity;connectionist language model;left-associative grammar;finite state machines;morphological parser;chart parsing;lexicalization;unification grammars;and swift answer algorithm

B0100; B6130; C4210; C4290; C7820; C1140Z; C6180N
computational linguistics;grammars;language translation;natural languages;probability;speech recognition
information based passing;disjunctive constraint satisfaction;head driven bidirectional passing;head driven parsing;probabilistic

passing;speech recognition;dependency grammar passing;combinatory grammars;Tomita algorithm;computational complexity;connectionist language model;left associative grammar;finite state machines; morphological parser;chart parsing;lexicalization;unification grammars;swift answer algorithm

C90003330
Indexing medical reports in a multimedia environment: the RIME experimental approach
12th Annual International ACMSIGIR Conference on Research and Development in Information Retrieval
Cambridge, MA, USA
25-28 June 1989
BERRUT C.; CHIARAMELLA Y.
Equipe Syst.Intelligents de Recherche d'Inf , L[illegible]AG, Grenoble, France
ACM
Conference paper
Economic
ENG
FR
SIGIR Forum (USA):SIGIR Forum
VOL. 23; NO. 1-2; spec issue.; PP. 187-97; 17 Ref.; OP. Fall 1988-Winter 1989
FASRDV
0163-5840
0-89791-321-3/89/0006-0187$1.50

Focuses on the RIME system aimed at the indexing of medical reports in a multimedia environment.This particular application is viewed to be appropriate for a large set of needs within large user communities: domain experts dealing with on-line specialized documentation such as software engineers, medical specialists and so on.In this application textual information appears as an interesting media to accessing related pictures in the data base.After presenting the application and a study of the particular corpus involved, the authors define a semantic model for the documents based on a conceptual language.They detail the indexing process and its various linguistic components, essential for the translation of medical reports

C7140; C7240
computational linguistics;indexing;language translation;medical administrative data processing;vocabulary
medical report indexing;multimedia environment;RIME experimental approach;large user communities;domain experts;on line specialized documentation;software engineers;medical specialists;textual information;data base;semantic model;documents;conceptual language; indexing process;linguistic components;translation

C90004601
Design of LMT: a Prolog-based machine translation system
MCCORD M C.
IBM Thomas, J.Watson Res.Center, Yorktown Heights, NY, USA
Journal paper
Practical
ENG
US
Comput.Linguist.(USA).Computational Linguistics
VOL. 15; NO. 1; PP. 33-52; 38 Ref.; OP. March 1989
CLINEE
0362-613X
0362-613X/89/010033-52$03.00

LMT (logic-based machine translation) is an experimental English-to-German MT system, being developed in the framework of logic programming.The English analysis uses a logic grammar formalism, modular logic grammar, which allows logic grammars to be more compact, and which has a modular treatment of syntax, lexicon, and semantics.The English grammar is written independently of the task of translation.LMT uses a syntax-to-syntax transfer method for translation, although the English syntactic analysis trees contain some results of semantic choices and show deep grammatical relations.Semantic type checking with Prolog inference is done during analysis and transfer The transfer algorithm uses logical variables and unification to good advantage:transfer works in a simple left-to-right, top-down way.After transfer, the German syntactic generation component produces a surface structure tree by application of a system of tree transformations.These transformations use an augmentation of Prolog pattern matching.LMT has a single lexicon, containing both source and transfer information, as well as some idiosyncratic target morphological information.There is a compact external format for this lexicon, with a lexical preprocessing system that applies defaults and compiles it into an internal format convenient for the syntactic components.During lexical preprocessing, English morphological analysis can be coupled with rules that synthesize new transfer entries

C7820
computational linguistics;language translation
semantic type checking;source information;LMT;Prolog based machine translation system;logic based machine translation;English to German MT system;logic programming;English analysis;logic grammar formalism;modular logic grammar;syntax;lexicon;English grammar; syntax to syntax transfer method;English syntactic analysis trees, Prolog inference;transfer algorithm;logical variables;unification, left to right;top down;German syntactic generation;surface structure tree;tree transformations;Prolog pattern matching; transfer information;idiosyncratic target morphological information;lexical preprocessing system;defaults;internal format; English morphological analysis;rules

C90004600
SPPS-the system of semi-automatic translation of professional text from English into Czech
STROSSA P.
Vysoka Skola Ekonomicka, Katedra Vedeckotech.Inf., Praha, Czechoslovakia
Journal paper
Practical
CZE
CS
Cesk.Inform.Teor.Praxe (Czechoslovakia);Ceskoslovenska Informatika, Teorie a Praxe
VOL. 31; NO. 8; PP. 231-5; 9 Ref.; OP. 1989
CITPBH
0322-8509

The article discusses the use of the interactive tool SPPS in the **computer**-**aided** **translation** of STI texts.This semi-automatic translation system is described in detail, its fundamental properties being presented.English language is required as the input and Czech language is produced as the output.SPPS's output requirements are discussed from the general linguistic point of view, along with the implementation software and corresponding data structures.Debugging of the system is described.Further developments of SPPS, along with its technical demands and applications, are presented

C7820
data structures;language translation
system debugging;professional text;interactive tool;SPPS;computer aided translation;STI texts;semi automatic translation system; English language;Czech language;implementation software;data structures

C89070243
Computerized text interpretation
MIRAM G. E.
Journal paper
Practical, Theoretical mathematical
ENG
ZZ
Nauchno-Tekh.Inf.2 (USSR):Nauchno-Tekhnicheskaya Informatsiya, Seriya 2
Autom.Doc.Math.Linguist.(USA);Automatic Documentation and Mathematical Linguistics
VOL. 22; NO. 9; PP. 19-23; 8 Ref.; OP. 1988
VOL. 22; NO. 5; PP. 32-9; OP. 1988
NIPSBP
ADMLAE
0548-0027
0005-1055
0005-1055/88/$20.00

Designers of automatic information processing systems are increasingly concerned with providing a formal conceptual representation of the problem universe.For systems operating with natural language texts, this means extracting from the natural language its conceptual content and endowing the content with a form, making it susceptible to computer input and processing according to a desired order.The author describes the results of a study of the semantic structure of English sc[illegible] texts, based on a model of distributional semantic classes (DSC) obtained with the SIMPAR-SMIT software package for the AVESTA national databank.The model and the analysis procedures are based on a novel application of the concept of distributional semantic analysis

C7820; C61300; C4290; C7300
computational linguistics;language translation;linguistics;natural languages,natural sciences computing;word processing
computerized text interpretation;interactive machine translation, automated unified thermophysics subscription system;automatic information processing systems;formal conceptual representation; problem universe;natural language texts;conceptual content; computer input;semantic structure;English science texts; distributional semantic classes;DSC;SIMPAR SMIT software package, AVESTA national databank;analysis procedures;distributional semantic analysis

C89064144
Qualitative and quantitative characteristics of Japanese sentences
ISHIZAKI S.; ISAHARA H.
Journal paper
Experimental
ENG
ZZ
Bull Electrotech.Lab.(Japan);Bulletin of the Electrotechnical Laboratory
VOL. 53; NO. 5; PP. 462-84; 6 Ref.; OP 1989
DESIA7
0366-9092

Reports on analysis of Japanese sentences conducted to determine their qualitative and quantitative characteristics which are considered to contribute to relative complexity of the Japanese language.Twenty-six textbooks used for classes of Japanese language, mathematics, science, and social study at elementary schools (specifically for second and fifth graders), junior high schools, and high schools were sampled to obtain some 2700 sentences for examination.The objects of quantitative analysis included the length of sentences, the number of verbs, adjectives, and adjectival verbs, and the numbers of modifying phrases and parallel structures.The qualitative examination encompassed homonyms, morphological ambiguities, sentence styles, parallel structures, ellipses, and anaphora.Five sentences were chosen from 16 textbooks (one each grade, one each subject).These selected sentences identified as those having the average characteristics are expected to serve as data against which present and future machine translation systems are evaluated

C7820; C6180N
computational linguistics;language translation;natural languages machine translation systems research;qualitative characteristics; language complexity;textbook samples;natural language processing; quantitative characteristics;Japanese sentences;analysis;Japanese

language;quantitative analysis;length of sentences;number of verbs; adjectives;adjectival verbs,numbers of modifying phrases;homonyms; morphological ambiguities;sentence styles;parallel structures; ellipses;anaphora

C89058988
On formalisms and analysis, generation and synthesis in machine translation
Fourth Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Manchester, UK
10-12 April 1989
YUSOFF Z.
Univ.Sains Malaysia, Penang, Malaysia
Conference paper
Practical
ENG
MY
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. xxv+326; PP. 319-26; 15 Ref.; DP. 1989
The author is interested in formalisms which are being used or have applications in the domain of machine translation (MT).His interest lies mainly in their role in the domain in terms of the ease in expressing linguistic knowledge required for MT, as well as the case of implementation in MT systems.He begins by discussing formalisms within the general context of MT, clearly separating the role of linguistic formalisms on one end, which are more apt for expression linguistic knowledge, and on the other, the SLLPS which are specifically designed for MT systems.He argues for another type of formalism, the general formalism, to bridge the gap between the two.Next he discusses the role of formalisms in analysis and in generation, and then more specific to MT, in synthesis.He sums up with the building of a compiler that generates a synthesis program in SLLP from a set of specifications written in a general formalism
C7820; C6180N; C4210
computational linguistics;formal specification;grammars;language translation
linguistic knowledge;linguistic formalisms;SLLPS;general formalism; compiler;synthesis program;specifications

C89058987
Situation semantics and machine translation
Fourth Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Manchester, UK
10-12 April 1989
RUPP C. J.
Manchester Univ.Inst.of Sci.& Technol., UK
Conference paper
Practical
ENG
GB
Assoc.Comput;Linguistics;Morristown, NJ, USA
NP. xxv+326; PP. 308-18; 20 Ref.; DP. 1989
The author demonstrates that the enriched theoretical vocabulary of situation semantics offers a more intuitive characterisation of the translation process, than was possible using more traditional semantic theories.This demonstration takes the form of a formalisation of the most commonly used method for MT in terms of situation semantic constructs.He considers what the theory of situation semantics has to offer to an MT application.The paper consists of a basic introduction to the machinery of situation semantics, an examination of the problem of translation, a formal description of a transfer-based MT system and some examples of the kind of lexical transfer one would expect to define in such a system
C7820; C6180N; C4210
computational linguistics;formal languages;language translation; natural languages
machine translation;theoretical vocabulary;situation semantics; formalisation;transfer based MT system;lexical transfer

C89058986
An approach to sentence-level anaphora in machine translation
Fourth Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Manchester, UK
10-12 April 1989
VAN NOORD G., DORREPAAL J.; ARNOLD D.; KRAUWER S., SADLER L. DES TOMBE L.
Foundation of Language Technol., State Univ.of Utrecht, Netherlands
Conference paper
Practical
ENG
NL
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. xxv+326; PP. 299-307; 14 Ref.; DP. 1989
Theoretical research in the area of machine translation usually involves the search for and creation of an appropriate formalism.An important issue in this respect is the way in which the compositionality of translation is to be defined.The authors introduce the anaphoric component of the Mimo formalism.It makes the definition and translation of anaphoric relations possible, relations which are usually problematic for systems that adhere to strict compositionality.In Mimo, the translation of anaphoric relations is compositional.The anaphoric component is used to define linguistic phenomena such as wh-movement, the passive and the binding of reflexives and pronouns monolingually.The actual working of the component is shown by means of a detailed discussion of wh-movement
C7820; C6180N
computational linguistics;language translation;natural languages
sentence level anaphora;machine translation;compositionality;Mimo formalism;linguistic phenomena;wh movement;passive;binding; reflexives;pronouns

C89058985
A new view on the process of translation
Fourth Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Manchester, UK
10-12 April 1989
BATEMAN J. A.; KASPER R. T.; SCHUTZ J. F. L.; STEINER E. H.
Inf.Sci.Inst., Univ.of Southern California, Marina del Rey, CA, USA
Conference paper
Practical
ENG
US
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. xxv+326; PP. 282-90; 26 Ref.; DP. 1989
The authors describe a framework for research into translation that draws on a combination of two existing and independently constructed technologies: an analysis component developed for German by the EUROTRA-D group of IAI and the generation component developed for English by the Penman group at ISI.They present some of the linguistic implications of the research and the promise it bears for furthering understanding of the translation process
C7820; C6180N
computational linguistics;language translation;natural languages
translation;analysis component;German;EUROTRA D;English;Penman

C89058984
Translation by structural correspondences
Fourth Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Manchester, UK
10-12 April 1989
KAPLAN R. M.; NETTER K.; WEDEKIND J.; ZAENEN A.
Xerox Palo Alto Res.Center, CA, USA
Conference paper
Practical
ENG
US
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. xxv+326, PP. 272-81; 18 Ref.; DP. 1989
The authors sketch and illustrate an approach to machine translation that exploits the potential of simultaneous correspondences between separate levels of linguistic representation, as formalized in the LFG notion of codescriptions.The approach is illustrated with examples from English, German and French where the source and the target language sentence show noteworthy differences in linguistic analysis
C7820; C6180N
computational linguistics;language translation;natural languages
structural correspondences;machine translation;linguistic representation;LFG notion;codescriptions;English;German;French; target language sentence;linguistic analysis

C89058983
A descriptive framework for translating speaker's meaning-towards a dialogue translation system between Japanese and English
Fourth Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Manchester, UK
10-12 April 1989
KUME M.; SATO G. K.; YOSHIMOTO K.
ATR Interpreting Telephony Res.Labs., Kyoto, Japan
Conference paper
Practical
ENG
JP
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. xxv+326; PP. 264-71; 13 Ref.; DP. 1989
A framework for translating speaker's meaning or intention is proposed based on two notions illocutionary force types (IFTs) for analysis and decision parameters (DPs) for generation.IFTs are a certain kind of classification of utterances concerning speaker's meaning.DPs present background information of language use in order to derive an appropriate expression from speaker's meaning.In Japanese, IFT's can be derived automatically through syntactical constraints.To generate appropriate expressions, language-specific communication strategies related to DP values should be given a priori.The whole process is performed in a unification-based framework
C7820; C6180N
computational linguistics;language translation;natural languages
MT system;descriptive framework;meaning;dialogue translation system;Japanese;English;intention;illocutionary force types; decision parameters;utterances;background information;syntactical constraints;unification based framework

C89057065
Interactive incremental chart parsing
Fourth Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Manchester, UK
10-12 April 1989
WIREN M.
Dept.of Comput.& Inf.Sci., Linkoping Univ., Sweden
Conference paper
Practical
ENG
SE

Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. xxv+326; PP. 241-8; 24 Ref.; DP. 1989
Presents an algorithm for incremental chart parsing, outlines how this could be embedded in an interactive parsing system, and discusses why this might be useful.Incremental parsing means that input is analysed in a piecemeal fashion, in particular allowing arbitrary changes of previous input without exhaustive reanalysis.Interactive parsing means that the analysis process is prompted immediately at the onset of new input, and possibly that the system then may interact with the user in order to resolve problems that occur.The combination of these techniques could be used as a parsing kernel for highly interactive and 'reactive' natural-language processors, such as parsers for dialogue systems, interactive **computer**-**aided** **translation** systems, and language-sensitive text editors.An incremental chart parser embodying the ideas put forward has been implemented, and an embedding of this in an interactive parsing system is near completion
C6180N; C4210
computational linguistics;grammars;interactive systems;natural languages
incremental chart parsing;interactive parsing system;piecemeal fashion;parsing kernel;natural language processors;dialogue systems;computer aided translation systems,language sensitive text editors

C89057055
The organization of the Rosetta grammars
Fourth Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Manchester, UK
10-12 April 1989
ODIJK J.
Philips Res.Labs., Eindhoven, Netherlands
Conference paper
Practical
ENG
NL
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. xxv+326; PP. 80-6; 7 Ref.; DP. 1989
The organization of the grammars in the Rosetta machine translation system is described and it is shown how this organization makes it possible to translate between words of different syntactic categories in a systematic way.It is also shown how the organization chosen makes it possible to translate small clauses into full clauses and vice versa.The central concept worked out here in some detail is the concept of partial isomorphy between subgrammars.The system as described has been implemented and is being tested
C6180N; C4210; C7820
computational linguistics,grammars,language translation,natural languages
Rosetta grammars;machine translation system;syntactic categories, small clauses;full clauses;partial isomorphy;subgrammars

C89057054
Ambiguity resolution in the DMTRANS PLUS
Fourth Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Manchester, UK
10-12 April 1989
KITANO H.; TOMABECHI H.; LEVIN L.
Carnegie Mellon Univ., Pittsburgh, PA, USA
Conference paper
Practical
ENG
US
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. xxv+326; PP. 72-9; 28 Ref.; DP. 1989
The authors present a cost-based (or energy-based) model of disambiguation.When a sentence is ambiguous, a parse with the least cost is chosen from among multiple hypotheses.Each hypothesis is assigned a cost which is added when: (1) a new instance is created to satisfy reference success, (2) links between instances are created or removed to satisfy constraints on concept sequences, and (3) a concept node with insufficient priming is used for further processing.This method of ambiguity resolution is implemented in DMTRANS PLUS, which is a second generation bi-directional English/Japanese machine translation system based on a massively parallel spreading activation paradigm
C6180N; C7820
computational linguistics;language translation;natural languages
natural language understanding;direct memory access;disambiguation; parse;least cost;multiple hypotheses;concept sequences;ambiguity resolution;DMTRANS PLUS;English Japanese machine translation system;massively parallel spreading activation

C89057050
Fourth Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Fourth Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Manchester, UK
10-12 April 1989
Conference proceedings
Practical
ENG
ZZ
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. xxv+326; DP. 1989
The following topics were dealt with: computational lexical semantics;parsing;grammars;natural language processing;expert systems;knowledge representation;logic programming;text-to-speech systems;intelligent tutors;knowledge acquisition;discourse representation;anaphora resolution;unification grammars;and machine translation
C6180N; C4210; C7820; C6170
computational linguistics;grammars;knowledge engineering;language translation;natural languages
computational lexical semantics;parsing;grammars;natural language processing;expert systems;knowledge representation;logic programming;text to speech systems;intelligent tutors,knowledge acquisition;discourse representation;anaphora resolution; unification grammars;machine translation

C89057040
Textual and computational linguistics
FERRARI G.
Journal paper
Practical
ITA
ZZ
Sist.Impresa (Italy);Sistemi & Impresa
VOL. 35; NO. 302; PP. 673-9; 24 Ref.; DP. April 1989
Reviews some of the major areas of computational linguistics and text processing.These include statistical linguistics, concordances and lemmatization (including some machine translation aspects), machine dictionaries, morphological analysis, text comprehension and generation, and stylistic analysis
C6180N; C7820; C61300
computational linguistics;language translation;linguistics;word processing
computational linguistics;text processing;statistical linguistics, concordances;lemmatization;machine translation;machine dictionaries;morphological analysis;text comprehension;stylistic analysis

C89058975
The potential of Swetra-a multilanguage MT system
SIGURD B.; GAWRONSKA WERNGREN B.
Dept.of Linguistics & Phonetics, Lund Univ., Sweden
Journal paper
Practical
ENG
SE
Comput.Transl.(USA);Computers and Translation
VOL. 3; NO. 3-4; PP. 237-50; 11 Ref.; DP. 1988-1989
COMTES
0884-0709
Swetra is a multilanguage MT system defined by the potentials of a formal grammar (standard referent grammar) and not by reference to a genre.Successful translation of sentences can be guaranteed if they are within a specified syntactic-format based on a specified lexicon.The authors discuss the consequences of this approach (grammatically restricted machine translation, GRMT) and describe the limits set by a standard choice of grammatical rules for sentences and clauses, noun phrases, verb phrases, sentence adverbials, etc.Such rules have been set up for English, Swedish and Russian, mainly on the basis of familiarity (frequency) and computer efficiency.However, restricting the grammar and making it suitable for several languages poses many problems for optimization.Sample texts-newspaper reports-illustrate the type of text that can be translated with reasonable success among Russian, English and Swedish
C7820; C4210; C4190
computational linguistics;grammars;language translation
machine translation;Swetra;multilanguage MT system;formal grammar, standard referent grammar;specified syntactic format;specified lexicon;grammatically restricted machine translation;GRMT; grammatical rules;sentences;clauses;noun phrases;verb phrases; sentence adverbials;English;Swedish;Russian;computer efficiency, newspaper reports

C89049801
Automatic computer recognition of German word types
XIE JINBAO; SUN JIEMING; WANG JIAN
Journal paper
Theoretical mathematical
CHI
ZZ
J.Shanghai Jiaotong Univ.(China),Journal of Shanghai Jiaotong University
VOL. 23; NO. 1; PP. 70-6; 5 Ref.; DP. 1989
SCTPDH
0253-9942
Word type recognition is the basis of natural language understanding and analysis.The authors give a description of the possibility of German word type recognition by computer according to the theory of pattern and pattern matching in SNOBOL as well as the flexibility of morphology in German.The software employed in German word type recognition has an accuracy better of over 95%
C6180N
computational linguistics;language translation,natural languages
automatic computer recognition,language pattern,language understanding;SNOBOL;German word type recognition

C89049800
Language and meaning
NAGAO M.
Fac.of Eng., Kyoto Univ., Japan
Journal paper
Theoretical mathematical
JAP
JP
J.Inst.Electron.Inf.Commun.Eng.(Japan),Journal of the Institute of

Electronics, Information and Communication Engineers
VOL. 71; NO. 11; PP. 1157-62; 3 Ref.; DP. Nov. 1988
DJTGEB
0913-5693
The relationship between language and its meaning is described, methods of defining meaning are outlined from the standpoint of computational linguistics, and the role of language in symbols and images is discussed.The topical areas include: (1) meaning of word;(2) meaning of paragraph;(3) meaning of sentence;(4) meaning in translation;and (5) role of language in meaning
C6180N; C4290
computational linguistics,language translation;natural languages
language;meaning;computational linguistics;symbols;images;word;paragraph;sentence;translation

C89041185
European Community policy on MT
MT Machine Translation Summit.Manuscripts and Program
Hakone, Kanagawa-ken, Japan
17-19 Sept. 1987
ROLLING L.
Conference paper
General
ENG
ZZ
Toshiba Corp;Kawasaki, Japan
NP. 159; PP. 97-8; 0 Ref.; DP. 1987
One of the roles of the EC Commission is to help the European Community to overcome the language barriers that are presently hampering its cultural and economic unification.The Commission is supporting basic research, developing new tools and resources and promoting their implementation through the creation of compatible, user-friendly infrastructures.On the research side, the EUROTRA programme aims at supplying not only with a modular MT system covering all European languages, but also with a valuable test bed for further research in computational linguistics.In the framework of the ESPRIT programme, several projects are aimed at the integration of voice recognition devices in industrial equipment, but the main research project is one that has undertaken a thorough analysis of seven European languages with a view to creating reliable lexical resources for use of both text and speech translation.The Commission also contributes to the development of standards for linguistic tools and resources, including lexical data banks, text and speech corpora and multilingual thesauri.Another major effort to the Commission has been its contribution to the development of the SYSTRAN system for a number of European language pairs
C0230; C7820
government policies;language translation
MT;European Community;language barriers;user friendly infrastructures;EUROTRA programme;modular MT system;European languages;computational linguistics;ESPRIT programme;voice recognition devices;industrial equipment;standards;linguistic tools;resources;lexical data banks;speech corpora;multilingual thesauri;SYSTRAN system;European language pairs

C89045071
Prospects in machine translation
MT Machine Translation Summit.Manuscripts and Program
Hakone, Kanagawa-ken, Japan
17-19 Sept. 1987
HUTCHINS W. J.
Univ.of East Anglia, Norwich, UK
Conference paper
General
ENG
GB
Toshiba Corp;Kawasaki, Japan
NP. 159; PP. 48-52; 0 Ref.; DP. 1987
Reviews the state of the art in MT systems, noting that no operational systems can produce good quality output without placing restrictions on input texts or involving human assistance.MT systems under development are based on the syntax-oriented approach of computational linguistics.AI approaches offer the scope for considerable improvements.The best prospects for future fully automated translation systems will be those combining traditional linguistics approaches and knowledge based approaches
C7820; C6170; C6180N
computational linguistics;knowledge based systems;language translation
machine translation;MT systems;human assistance;syntax oriented approach;computational linguistics;AI approaches;fully automated translation systems;knowledge based approaches

C89041183
Governmental views of MT
MT Machine Translation Summit,Manuscripts and Program
Hakone, Kanagawa-ken, Japan
17-19 Sept. 1987
CZERMAK J. M.
Conference paper
General
ENG
ZZ
Toshiba Corp;Kawasaki, Japan
NP. 159; PP. 38-9; 0 Ref.; DP. 1987
Language will play an important role in future R&D work.It is amenable to computer representation.Language R&D still leaves many questions unresolved.Translation is one of these.The computer will contribute to the resolution of these questions.The discipline of computational linguistics may not be sufficiently mature.Computational linguistics will become a major field of international R&D cooperation
C0230; C7820
government policies;language translation
government views;language;MT;R D work;computer representation;computational linguistics

C89041069
Computer **aided** **translation** system and computerized dictionary
MT Machine Translation Summit.Manuscripts and Program
Hakone, Kanagawa-ken, Japan
17-19 Sept. 1987
YAMAOKA Y.
Conference paper
Practical
ENG
ZZ
Toshiba Corp;Kawasaki, Japan
NP. 159; PP. 141-2; 0 Ref ; DP. 1987
ISS Inc., are engaged in language services including translation services and are one of the leading companies in Japan in this field.Their customers require the highest possible quality of translation from English to Japanese for their sales materials, manuals, catalogues and so on.ISS have therefore introduced a **computer** **aided** **translation** system developed by Toshiba.They have been utilizing and improving this system to assist translation activities conducted by experienced and professional translators, mainly for the following purposes: standardisation and unification of terminology and customization of terminology
C7820
language translation;linguistics;standards
ISS;computerized dictionary;language services;translation services;leading companies;Japan;quality;English;Japanese;sales materials;manuals;catalogues;computer aided translation system;Toshiba, translation activities;professional translators;standardisation, unification;customization;terminology

C89041037
Fujitsu Machine translation system
MT Machine Translation Summit.Manuscripts and Program
Hakone, Kanagawa-ken, Japan
17-19 Sept. 1987
UCHIDA H.
Fujitsu Labs.Ltd., Japan
Conference paper
Practical; Product reviews
ENG
JP
Toshiba Corp;Kawasaki, Japan
NP. 159; PP. 129-34; 0 Ref.; DP. 1987
Due to the rapid advancement of both computer technology and linguistic theory, machine translation systems are coming into practical use.Fujitsu has two machine translation systems.ATLAS-I is a syntax-based machine translation system.ATLAS II is a semantic-based system which aims at high quality multilingual translation.The ATLAS II translation mechanism is explained;it involves analysis, transfer and generation processes
C7820; C61300
computational linguistics;language translation;linguistics;word processing
transfer process;analysis process;Japanese English translation, Fujitsu Machine translation;computer technology;linguistic theory;machine translation systems;ATLAS II;semantic based system;high quality multilingual translation;translation mechanism;generation processes

C89041060
The need for MT in Indonesia
MT Machine Translation Summit.Manuscripts and Program
Hakone, Kanagawa-ken, Japan
17-19 Sept. 1987
SUDARWO I.
Conference paper
Practical
ENG
ZZ
Toshiba Corp;Kawasaki, Japan
NP. 159; PP. 113; 0 Ref.; DP. 1987
Summary form only given.Machine translation technology will bring new opportunities for Indonesia.An Indonesian agency is therefore conducting research and developing a prototype English-Indonesian **Computer** **aided** **translation** system.Involvement in a Japanese R&D project to develop multilanguage machine translation is also outlined
C7820; C0230
economic and sociologic effects;language translation;research and development management
MT;Indonesian agency;English Indonesian Computer aided translation system;Japanese R D project;multilanguage machine translation

C89036044
Kana-to-Kanji translation based on collocational analysis for non-segmented input
YAMASHINA M.; OBASHI F.
Human Interface Labs., NTT, Tokyo, Japan
Journal paper
Practical
ENG
JP
Rev.Electr.Commun.Lab.(Japan);Review of the Electrical Communication Laboratories

VOL. 37; NO. 1; PP. 65-70; 13 Ref.; DP. Jan. 1989
RELTAN
0029-067X
Proposes a new disambiguation method for Kana-to-Kanji translation.This method evaluates candidate sentences by measuring the number of word cooccurrence patterns (WCP) included in the candidate sentences.An automatic WCP extraction method is also developed and about 305000 sets of WCP are collected from example sentences in dictionaries by this method.Using a WCP matrix organized by semantic category, the mean number of candidate sentences in Kana-to-Kanji translation is reduced to about 1/10 of those produced by existing morphological methods, and results in a translation accuracy of 95%
C7820; C4290; C6180N; C6170
computational linguistics;knowledge based systems;language translation;natural languages
nonsegmented input;Kana to Kanji translation;collocational analysis;disambiguation method;candidate sentences;word cooccurrence patterns;semantic category;morphological methods

C89034563
Interdoc (indexing retrieval aid)
JOSCELYNE A.
Journal paper
Practical
RUS
ZZ
Lang.Technol.(Netherlands);Language Technology
NO. 11; PP. 28-31; 0 Ref.; DP. Jan.-Feb. 1989
LANTEB
The paper discusses the sophisticated documentation tool Interdoc, an indexing-retrieval enhancement to CAT (**computer** **aided** **translation**).Whereas CAT was primarily aimed at the dictionary-using translator, Interdoc is specifically designed to be used by a wide range of professionals, including corporate indexers and target-language end users.The core idea is that of the corporate knowledge system
C7250; C7240
indexing;information retrieval systems
indexing retrieval aid;documentation tool;Interdoc;CAT;computer aided translation;professionals;corporate indexers;target language end users;corporate knowledge system

C89032283
A mathematical model for translations of natural languages
KATZ E.; LEIFMAN L. J.; MARTY R. H.; ROBINSON S. M.
Dept.of Math., Cleveland State Univ., OH, USA
Journal paper
Theoretical mathematical
ENG
US
Inf.Sci.(USA);Information Sciences
VOL. 47; NO. 1; PP. 35-45; 10 Ref.; DP. Feb. 1989
ISIJBC
0020-0255
0020-0255/89/$03.50
Several mathematical models have been introduced in linguistics and in translations of languages.The essential mathematical tools used have been algebra, probability, logic, etc.The authors introduce a topological model for languages and their translations.Using this model, they prove that every text in a major language has a best approximation text in any other major language.They prove, similarly, that every text in any major language has a best approximation within the language itself.This permits the theoretical possibility of automatic translations
C4290; C7820
computational linguistics;language translation;natural languages
mathematical model;natural languages;linguistics;mathematical tools;algebra;probability;logic;topological model;approximation text;automatic translations

C89029779
CATEC-a **computer** **aided** **translation** of English to Chinese system
1988 International Conference on Computer Processing of Chinese and Oriental Languages.Proceedings
Toronto, Ont., Canada
29 Aug.-1 Sept. 1988
TOU J. T.
Center for Inf.Res., Florida Univ., Gainesville, FL, USA
Chinese Language Comput.Soc.;Chinese Canadian Inf.Processing Professionals;Philips Electron
Conference paper
Practical; Theoretical mathematical
ENG
US
Concordia Univ;Montreal, Que., Canada
NP. xvii+645; PP. 475-9; 15 Ref.; DP. 1988
Represents a new system for **computer**-**aided** **translation** of technical and scientific publications in English into Chinese language.This system is based upon the innovative idea of linguistic canonical transformation in order to incorporate the cultural aspects of a natural language.Via paraphrasing by computer, the computer, the messages and information contained in a complex sentence or a set of sentences are expressed in terms of several simple sentences.By making use of a knowledge-base of skilled translators' expertise, these sentences are converted to 'Chinese English' sentences which are referred to as linguistic canonical forms.The Chinese English sentences are then translated into Chinese text
C7820; C4210; C6170
knowledge based systems;language translation
computer aided translation;scientific publications;linguistic canonical transformation;cultural aspects;natural language;paraphrasing;complex sentence;knowledge base;linguistic canonical forms;Chinese English sentences

C89029766
The Saarbrucken Translation Service STS-**computer**-**aided** **translation** for specialised information centers
LUCKHARDT H. D.; ZIMMERMANN H. H.
Univ.des Saarlandes, Saarbrucken, West Germany
Journal paper
Practical
GER
DE
Nachr.Dok.(West Germany);Nachrichten fur Dokumentation
VOL. 39; NO. 6; PP. 351-6, 16 Ref.; DP. Dec. 1988
NADOAW
0027-7436
0027-7436/88/0612-0351$02.50/0
The paper presents the Saarbrucken **Computer**-**Aided** **Translation** Service (STS) being developed in the project MARIS (Multilingual Application of Reference-Oriented Information Systems) at the Information Science Department of the University of Saarbrucken.Intellectual and machine translation (esp.German to English) are combined in a joint system surrounding (translator's workstation),MARIS applies methods and (sub)systems developed for machine translation to titles, abstracts, and descriptors from German databases,About 2 million words have been translated so far.The MARIS project is funded by the Federal Ministry of Science and Technology
C7820; C7210
bibliographic systems,information services,language translation
Saarbrucken Translation Service;Saarbrucken Computer Aided Translation Service;MARIS;Multilingual Application of Reference Oriented Information Systems;German to English;titles;abstracts;descriptors;German databases

C89029765
An English-Indonesian **computer** **aided** **translation** system
SUKMADJAJA D.; SHIMURA M.
Agency for the Assessment & Appl.of Technol., Jakarta, Indonesia
Journal paper
Practical; Product reviews
ENG
ID
J.Jpn.Soc.Artif.Intell.(Japan);Journal of Japanese Society for Artificial Intelligence
VOL. 3; NO. 1; PP. 103-7; 6 Ref.; DP. Jan. 1988
JCGAED
0912-8085
This paper presents a sophisticated method for an English-Indonesian machine translation system called EICATS (English-Indonesian **computer** **aided** **translation** system).In general, a machine translation system consists of three main processes, namely analysis, transfer and generation.Depending on the method that is used in the transfer level, machine translation systems can be classified into four methods: syntactical direct, transfer, integration and the interlingua or pivot method.In EICATS, the analysis, transfer and generation processes are not handled as independent processes, but are integrated.Consequently, the translation process is done in real time, approaching the behaviour of a human translator model
C7820; C6170
expert systems;language translation
syntactical direct translation;transfer translation,integration translation;interlingua method;English Indonesian computer aided translation system;pivot method;analysis;generation

C89011017
Coherent analysis of argumentative discourse
Proceedings COMPSAC 88; The Twelfth International Computer Software and Applications Conference (Cat.No.88CH2611-2)
Chicago, IL, USA
5-7 Oct. 1988
COHEN R.
Dept.of Comput.Sci., Waterloo Univ., Ont., Canada
IEEE
Conference paper
Theoretical mathematical
ENG
CA
IEEE Comput.Soc.Press;Washington, DC, USA
NP. xvi+531; PP. 468-9; 2 Ref.; DP. 1988
0-8186-0873-0
CH2611-2/88/0000-0468$01.00
R.Cohen (Computational Linguistics, vol.13, no.1-2, p.11-24, 1987) has described a model for the analysis of arguments that includes: (1) a theory of expected coherent structure, which is used to limit analysis to the reconstruction of particular transmission forms;(2) a theory of linguistic clues which assigns a functional interpretation to special words and phrases used by the speaker to indicate the structure of the argument;and (3) a theory of evidence relationships which includes the demand for pragmatic analysis to accommodate beliefs not currently held.The author summarizes the prescriptions for coherent analysis, with a view to their application in the translation of technical material
C1230; C6180N
artificial intelligence;computational linguistics;language translation;natural languages
argumentative discourse;expected coherent structure;linguistic clues;functional interpretation;words;phrases;evidence relationships;pragmatic analysis;coherent analysis;translation, technical material

C8806C192
A dependency-based analysis of English for the purpose of machine translation.2nd edition
KIRSCHNER Z.
Book
ENG
ZZ
Univ.Karlovy;Prague, Czechoslovakia
NP. 157; DP. 1988
The book presents a description of the analysis of English in the framework of machine translation experimentation carried out by the linguistic group at Charles University in Prague.The project in question, called APAC2, represents the second experiment in a series of three.The book covers formal representation, program structure, morphemic analysis and dictionaries, the noun syntax, and the verb
C7820
computational linguistics;language translation,linguistics;natural languages
linguistics;natural languages;dependency based English analysis; English;machine translation;Charles University;Prague;APAC2;formal representation;program structure;morphemic analysis;dictionaries; noun syntax;verb

C88054486
Microcomputers technology transfer: case of developing multilingual software (English to Arabic conversion)
Technology Management 1.Proceedings of the First International Conference
Miami, FL, USA
17-19 Feb. 1988
HOSNI Y A.; KHALIL T. M.(Ed.); BAYRAKTAR B. A.(Ed.), EDOSOMWAN J. A.(Ed.)
Central Florida Univ., Orlando, FL, USA
Conference paper
Practical
ENG
US
Inderscience Enterprises;Geneva, Switzerland
NP. xxii+866; PP. 416-23; 5 Ref.; DP. 1988
0-907776-13-2
One of the most important problems facing the transfer of microcomputing technology is that most of the software is developed for English language speaking communities.This is mainly because the hardware is 'Latin' based and most of the programming languages are 'like English' languages.However, technological developments have resulted in the production of multilingual microcomputer hardware.Concentration now is in the development of 'foreign' software.The author presents the initial results of a research project dealing with transforming software written for English language users to Arabic language users with multilingual hardware.A computer program is presented which may facilitate the transformation process along with a sample application for a transformed project management program
C7820
computational linguistics;language translation,microcomputer applications;software packages
multi lingual software;Latin based,microcomputing technology, English language speaking communities;programming languages, technological developments;multilingual microcomputer hardware; research project;English language users;Arabic language users; computer program;transformed project management program

C88054412
On generating a German lexicon
RUPLIN F.; RUSSELL J.
State Univ.of New York, Stony Brook, NY, USA
Journal paper
Practical
ENG
US
CALICO J.(USA);CALICO Journal
VOL. 5; NO. 4; PP. 51-62; 0 Ref.; DP. June 1988
CALJEB
0742-7778
Building on a method of compressing lexical information, the authors have defined a set of algorithms providing the minimum amount of information necessary to generate all forms in the German lexicon and to detect spelling errors.Master stems were marked for part of speech and desinences with a view to also allowing possible inference drills for use in foreign language instruction
C7810C; C7820
computational linguistics;computer aided instruction;language translation;linguistics;teaching
master stems;compression;morphology;lexical information;algorithms, German lexicon;spelling errors;desinences;possible inference drills;foreign language instruction

C88039527
GPSG and German word order
Natural Language Parsing and Linguistic Theories
Friedenweiler, West Germany
April 1986
HAUENSCHILD C.; REYLE U.(Ed.); ROHRER C.(Ed.)
Tech.Univ.Berlin, West Germany
Conference paper
Practical; Theoretical mathematical
ENG
DE
Reidel;Dordrecht, Netherlands
NP. vi+30; PP. 411-31; 19 Ref.; DP 1988
The author's main concern is raising questions rather than giving answers.His starting point is Hans Uszkoreit's revised version of the LP (linear precedence) component (cf.Uszkoreit 1984 and 1986) within the GPSG formalism (cf.Gazdar et al.1985).He discusses some problems of Uszkoreit's approach that result from the fact that the whole complex phenomenon of German word order is described at a unique level of linguistic representation.He then proposes a somewhat speculative solution to some of these problems, which is based on a multi-level approach to analysis and generation within the context of machine translation
C4210; C4290; C7820
computational linguistics;grammars;language translation; linguistics;natural languages
LP component;GPSG;German word order;linear precedence;linguistic representation;machine translation

C88042934
A feature-based categorial morpho-syntax for Japanese
Natural Language Parsing and Linguistic Theories
Friedenweiler, West Germany
April 1986
WHITELOCK P. J.; REYLE U.(Ed.); ROHRER C.(Ed.)
Dept.of Artificial Intelligence, Edinburgh Univ., UK
Conference paper
Practical; Theoretical mathematical
ENG
GB
Reidel;Dordrecht, Netherlands
NP. vi+30; PP. 230-61; 17 Ref.; DP. 1988
Describes an experiment to investigate the characterisation of Japanese morpho-syntax within a lexicalist framework.It forms part of a study into English and Japanese grammars from the parochial, contrastive and universal viewpoints, which is intended to support the implementation of machine translation systems between the two languages
C7820; C4210; C4290
computational linguistics;grammars;language translation; linguistics;natural languages
feature based categorial morpho syntax;Japanese;lexicalist framework;English;Japanese grammars;machine translation systems

C88039514
Natural Language Parsing and Linguistic Theories
Natural Language Parsing and Linguistic Theories
Friedenweiler, West Germany
April 1986
REYLE U.(Ed.); ROHRER C.(Ed.)
Conference proceedings
Practical; Theoretical mathematical
ENG
ZZ
Reidel;Dordrecht, Netherlands
NP. vi+30; DP. 1988
The following topics were dealt with: separating linguistic analyses from linguistic theories;applicability of indexed grammars to natural languages;natural language toolkit;extension of LR-parsing for lexical functional grammar;efficiency-oriented LFG parser,parsing with a GB-grammar;combining categorial grammar and unification;feature-based categorial morpho-syntax for Japanese;treatment of the French adjectif detache in lexical functional grammar;problems of coordination in German;German word order and universal grammar;nonlocal-dependencies and infinitival constructions in German;GPSG and German word order;nested Cooper storage: proper treatment of quantification in ordinary-noun phrases;and compositional semantics for LFG.Abstracts of individual papers can be found under the relevant classification codes in this or other issues
C4210; C7820; C4290
computational linguistics;grammars;language translation; linguistics;natural languages
linguistic analyses;linguistic theories;indexed grammars;natural language toolkit;LR parsing;lexical functional grammar;efficiency oriented LFG parser;GB grammar;categorial grammar;unification; feature based categorial morpho syntax;Japanese;French;adjectif detache;coordination;German word order;universal grammar;nonlocal dependencies;infinitival constructions;GPSG;nested Cooper storage; quantification;ordinary noun phrases;compositional semantics;LFG

C88042922
The aims of the French national project of **computer**-**aided** **translation**
Workshop on Industrial Product Documentation
Marienamn, Finland
15-18 June 1987
PECCOUD F.
Groupe d'Etude pour la Traduction Autom., Univ.Sci., Tech.et Med.de Grenoble, France
Conference paper
Practical; Product reviews
ENG
FR
Int.Forum Inf.Doc.(USSR)
VOL. 13; NO. 1; PP. 11-13; 0 Ref.; DP. Jan. 1988
IFIDO7
0304-9701
The French national project of **computer**-**aided** **translation** (Traduction Assistee par Ordinatour, TAO) has led to the implementation of a production system called CALLIOPE-AERO using the software tool ARIANE.This system permits automatic translation into English of texts written in French in the field of aircraft maintenance.After a brief account of the architecture of the system, the author indicates its main performance characteristics as measured in that application and then considers what economic conclusions should be drawn from this first

full-scale experience regarding the development of a linguistic software industry for **computer**-**aided** **translation**
C7820; C7460
aerospace computing;language translation
avionics manuals;French national project;computer aided translation;Traduction Assistee par Ordinateur;TAO;CALLIOPE AERO; software tool;ARIANE;automatic translation;aircraft maintenance; performance characteristics;linguistic software industry

C88041637
Workshop on Industrial Product Documentation
Mariehamn, Finland
15-18 June 1987
Conference proceedings
General; Practical
ENG
ZZ
Int.Forum Inf.Doc.(USSR)
VOL. 13; NO. 1; DP. Jan. 1988
IFID07
0304-9701
The following topics were dealt with: the conference aimed to examine the whole flow of information from industrial producers to those who use the industrial products.It included current practices and recent development concerning documentation for offshore platforms, building industry, aviation, and electronics industry.Topics covered included: the aims of the French national project of **computer** **aided** **translation**, and some holistic and sociodynamic aspects of industrial product documentation.Abstracts of individual papers can be found under the relevant classification codes in this or other issues
C7200; C7820
information science;technical presentation;user manuals
offshore platforms;building industry;aviation;electronics industry; computer aided translation;industrial product documentation

C88042920
A holography-based **computer**-**aided** **translation** system-conceptual analysis
PRIMOV G. V.
Ind.AI Syst., Microprocessor Syst.Inst., Sofia, Bulgaria
Journal paper
Practical
ENG
BG
Cybernetica (Belgium)
VOL. 31; NO. 1; PP. 43-55; 30 Ref.; DP. 1988
CYBEAS
0011-4227
The paper aims at presenting a conceptual analysis of a new **computer**-**aided** **translation** system (CATS) paradigm, taking into consideration basic human information processing capabilities.Another feature of the suggested approach is that it allows principal hardware implementation of different input text analysis phases, thus eliminating the need of the large, complicated resident software used for language parsing, aiming at radical CATS architecture changes The idea is to overcome extant hardware limitations by using the advantages of parallel access and associative information processing in holographic storage media
C7820
holographic storage;language translation
holography based computer aided translation system;conceptual analysis;human information processing;text analysis,language parsing;parallel access;associative information processing; holographic storage

C88026474
APAC3-2; an English-to-Czech machine translation system
KIRSCHNER Z.
Book
Practical; Theoretical mathematical
ENG
ZZ
Univ Karlovy;Prague, Czechoslovakia
NP, vii+182; 14 Ref.; DP. 1987
Looks at an English-to-Czech machine translation system called APAC3-2.The text is divided into three chapters, the first of which presents a brief survey of the theoretical linguistic background and of [illegible] ftware tools used.Chapter 2 is devoted to a discussion [illegible] features of the system, of its limits, of the (part [illegible] ative) representation of factual knowledge, [illegible] now ambiguities are being solved, of the fail-soft [illegible] used, and so on.A detailed description of the present shape of the program is contained in Chapter 3
C7820; C4290
computational linguistics;language translation
linguistics;English;Czech;machine translation;APAC3 2;software tools;factual knowledge

C88026463
Text translation
THWAITE D.
Alcatel Co., Harlow, UK
Journal paper
General; Practical
ENG
GB
Comput.Syst.(UK)
VOL. 8; NO. 1; PP. 22-3; 0 Ref.; DP. Jan. 1988
COSYEG
0264-4193
Computer **aided** **translation** has been around for many years, but how practical is it and how can it be best used?The author investigates the latest developments .
C7820
language translation
text translation;computer aided translation

C88021476
A Korean-English machine translation system based on lexical association grammar
Proceedings of TENCON 87: 1987 IEEE Region 10 Conference 'Computers and Communications Technology Toward 2000' (Cat.No.87CH2423-2)
Seoul, South Korea
25-28 Aug. 1987
LEE J. K.; HAN S. K.; LEE S. H.
Dept.of Electron.Eng., Inha Univ., Incheon, South Korea
IEEE;Korea Inst.Electron.Eng.;Minist.Commun.;et al
Conference paper
Practical
ENG
KR
IEEE;New York, NY, USA
NP. 3 vol. 1380; PP. 516-19 vol.2; 7 Ref.; DP. 1987
CH2423-2/87/0000-0516$01.00
The implementation of a Korean-English machine translation system is described.To solve the complexity of machine translation problems effectively, a lexical association grammar (LAG) is proposed that is based on the analysis of a language cognition system.As LAG has both universality to represent the conceptual structure of language and particularly to generate the surface structure, it is especially effective for translation between different language families.The translation process is explained in detail
C7820; C4210; C4290
computational linguistics;grammars;language translation
Korean English machine translation system;lexical association grammar;LAG;language cognition system;universality;conceptual structure;particularly;surface structure

C88015568
Frames on nouns
STEINER E.; WINTER THIELEN J.
IAI, Saarbrucken, West Germany
Journal paper
Practical
ENG
DE
Sprache Datenverarb.(West Germany)
VOL. 11; NO. 1; PP. 29-35; 16 Ref.; DP. 1987
SPDADH
0343-5202
This paper discusses some central issues of the questions of the semantic valency of nouns in the context of the multilingual machine translation project Eurotra.It discusses some of the most influential positions which are known from the linguistic literature.It outlines approaches to non de-verbal nouns and to de-verbal nouns which are not identical, but within one overall theoretical perspective.The treatment of semantic elements which are not valency bound is discussed.Interesting questions connected with an implementation of the framework are examined as well as areas of research and implementation which are necessary
C7820
computational linguistics;language translation;linguistics
semantic valency;nouns;machine translation;Eurotra

C88015567
The dictionary in the Eurotra engineering framework
MAAS H. D.
IAI, Saarbrucken, West Germany
Journal paper
Practical
ENG
DE
Sprache Datenverarb.(West Germany)
VOL. 11; NO. 1; PP. 15-21; 0 Ref.; DP. 1987
SPDADH
0343-5202
The author gives an overview of the lexicon framework as worked out by the Dictionary Task Force of Eurotra.The article reflects the state of the art by the end of March 1987.It covers what generators do;the place of the inner dictionary in a generator;treatment of frames;treatment of idioms;structure-to-feature translation;dictionary coding;relations between dictionaries of different levels;and examples for using the dictionary in analysis and in generation
C7820
computational linguistics;glossaries;language translation; linguistics
machine translation;lexicon framework;Eurotra;generators;frames; idioms;dictionary coding;dictionaries

C87062783
A software tool for research in linguistics and lexicography: application to Somali
LECARME J.; MAURY C.
CNRS, Valbonne-Sophia Antipolis, France
Journal paper
Practical
ENG
FR
Comput.& Transl.(USA)
VOL. 2; NO. 1; PP. 21-36; 27 Ref.; DP. Jan.-March 1987

COMTES
0884-0709

General-purpose database management systems, whose structure is built in, are not an appropriate solution to situations where problems of translation or areas of research cannot be bounded in advance, for example, when lexicography and linguistic research are closely related.Consequently, an original system has been developed, and is being applied to linguistic and lexicographical data on the Somali language.The collaborative project led to the creation of an automatic lexicogrammatical data management system.The basic hardware provided for this application was a Plessey LSI-11/03 microcomputer with a 64-Kbyte memory, two 1-Mbyte floppy disk drives, a video terminal, and a printer.Although originally designed for the Somali language, the system is independent of any language and of any specific application.First the authors present the linguistic context and the computer context.Then they describe the system itself, with special emphasis on the original aspects.Finally some examples of work sessions are presented

C7820; C6160
database management systems;languaga translation;linguistics
computer aided translation;linguistics;lexicography;database
management syetems;translation;lexicography;linguistic research;
Somali language;automatic lexicogrammatical data management system;
Plessey LSI 11 03 microcomputer;64 Kbyte memory

C87056423
Text processing in the Leningrad research group 'Speech Statistics'-theory, results, outlook
PIOTROWSKI R.
Herzen Pedagogical Inst., Leningrad, USSR
Journal paper
General
ENG
SU
Lit.& Linguist.Comput.(GB)
VOL. 1; NO. 1; PP. 36-40; 19 Ref.; DP. 1986
0268-1145

The article discusses some semiotic informational aspects of language along with their interpretation in terms of computational linguistics.The paper describes and indicates the resolution of some semiotic and linguistic paradoxes which create, at present, a rejecting barrier between natural language and the computer

C7820
computational linguistics;language translation
text processing;machine translation;semiotic informational aspects;
linguistic paradoxes;natural language

C87044316
'SPSS'-an algorithm and data structures design for a machine aided English-to-Czech translation system
STROSSA P.
Journal paper
Practical; Theoretical mathematical
ENG
ZZ
Prague Bull.Math.Linguist.(Czechoslovakia)
NO. 47; PP. 25-36; 4 Ref.; DP. 1987
PBMLAT
0032-6585

Some of the problems caused by ambiguity in machine translation systems are outlined.Approaches to system design are discussed.The author describes a system which offers alternative translations of ambiguous words and phrases to a user.The task is examined from the linguistic point of view.The data structures and principles of the algorithm are presented

C7820; C4290
computational linguistics;data structures;language translation
linguistics;SPSS;machine aided English to Czech translation system;
ambiguity;machine translation;alternative translations;data
structures;algorithm

C87044315
Information systems as translators' aids
KAZAKEVICH O. A.
Journal paper
Application
ENG
ZZ
Nauchno-Tekh.Inf.Ser.2 (USSR)
Autom.Doc.& Math.Linguist.(USA)
VOL. 20; NO. 7; PP. 1-6; 23 Ref.; DP. 1986
VOL. 20; NO. 4; PP. 1-10; DP. 1986
NIPSBP
ADMLAE
0548-0027
0005-1055
0005-1055/86/$20.00

Research and development in computerized translation is conducted in two areas: **computer**-**aided** **translation** performed by human translators (using computerized terminology dictionaries and data banks) and automatic translation performed by computers assisted by human translators.There has been a growing realization that fully automatic translation can be achieved only in exceptional cases where a very general idea of the contents of the original would be sufficient for the user, or if the original texts have a simple and standard format.In all the other situations, researchers are becoming increasingly aware that translation programs have to rely on human participation at certain points in the process

C7820
language translation
information systems;computerized translation;computer aided
translation;terminology dictionaries;automatic translation;human
translators;standard format

C87039936
Features of processing of unidentified words in a machine translation system
KOROSTELEV L. Y.
Journal paper
Practical; Theoretical mathematical
ENG
ZZ
Nauchno-Tekh.Inf.Ser.2 (USSR)
Autom.Doc.& Math.Linguist.(USA)
VOL. 19; NO. 2; PP. 23-8; 16 Ref.; DP. 1985
VOL. 19; NO. 4; PP. 95-102; DP. 1985
NIPSBP
ADMLAE
0548-0027
0005-1055
0005-1055/86/$20.00

An efficient mechanism for recognizing grammatical characteristics of keywords has been developed for machine translation and other automatic text-processing systems.By matching parts of words (mostly word endings) with similar parts of words sorted in the system, the syntactic category and some of the grammatical (and also semantic) characteristics are determined.Efficient operation of this algorithm makes it possible to continue with the analysis, but it does not produce the translation of the partially identified word.The new component of the unidentified word processing algorithm has two tasks: separating the group of unidentified words by dividing it into new words and misspelled words (diagnosis of distortions) and reconstructing the prototypes of misspelled words (distortion correction).Both tasks are accompanied with a common strategy based on a formal description of the word's graphematics, i.e.regularities in the combinations of graphemes in words

C4290; C7820
computational linguistics;language translation;natural languages;
spelling aids
computational linguistics;distortion diagnosis;unidentified word
processing;natural language processing;semantic characteristics;
machine translation system;grammatical characteristics;keywords;
automatic text processing systems;word endings;syntactic category;
partially identified word;misspelled words;distortion correction;
graphematics;graphemes

C87036531
Variables and grammatical categories in an Ariane model
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
GUILBAUD J. P.
GETA, Univ.Sci.et Med.de Grenoble, Saint-Martin d'Heres, France
Conference paper
Practical
FRE
FR
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 405-7; 4 Ref.; DP. 1986

Grammatical categories used in a translation model.Ariane are formalised, and the variable of the metalanguage used to describe the source and target languages of the model are discussed.Variables in the linguistic structure interface, SLI and other grammatical categories are considered under seven headings.Particular attention is given to the structure, morphology and syntax of the German language in this context

C7820; C4210
computational linguistics,grammars,language translation;natural
languages
translation;grammatical categories;Ariane model;metalanguage;
linguistic structure interface;SLI;morphology;syntax;German
language

C87033360
Strategies and heuristics in the analysis of a natural language in machine translation
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
YUSOFF Z.
Groupe d'Etudes pour la Traduction Autom., Grenoble Univ.,
Saint-Martin-d'Heres, France
Conference paper
Theoretical mathematical
ENG
FR
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 136-9; 12 Ref.; DP. 1986

The analysis phase in an indirect, transfer and global approach to machine translation is studied.The analysis conducted can be described as exhaustive (meaning with backtracking), depth-first and strategically and heuristically driven, while the grammar used is an augmented context free grammar.The problem areas, being pattern matching, ambiguities, forward propagation, checking for correctness and backtracking, are highlighted.Established results found in the literature are employed whenever adaptable, while suggestions are given otherwise

C4290; C7820
computational linguistics;context free grammars;language
translation;natural languages
natural language;machine translation;backtracking;augmented

context free grammar:pattern matching:ambiguities:forward propagation

C87036515
The translation method of Rosetta
LEERMAKERS R.; ROUS J.
Philips Res.Labs., Eindhoven, Netherlands
Journal paper
Practical
ENG
NL
Comput.& Transl.(USA)
VOL. 1; NO. 3; PP. [illegible]; 9 Ref.; DP. July-Sept. 1986
COMTE5
0884-0709
The authors explain and motivate the translation method of the Rosetta project.They present a stepwise way of unraveling the various aspects of machine translation.The general strategy is to view a translation system as being composed of an analysis part and a generation part connected by a transfer module, and to break down the latter in a systematic way.They do this by repeatedly identifying tasks inside the transfer module, which can be moved as new modules, with well-defined interfaces, into the (initially empty) analysis and generation parts.Thus, by each such move analysis and generation are augmented with a deeper level in a clear way, and the tranafer task is reduced accordingly
C7820; C4210
computational linguistics:grammars:language translation:natural languages
grammars:transfer complexity:Rosetta:translation method:machine translation:analysis part:generation part:transfer module

C87036514
Symmetric rules for translation of English and Chinese
WANYING JIN; SIMMONS R. F.
Dept.of Comput.Sci., Shanghai Univ.of Technol., China
Journal paper
Practical
ENG
CN
Comput.& Transl.(USA)
VOL. 1; NO. 3; PP. 153-67; 7 Ref.; DP. July-Sept. 1986
COMTE5
0884-0709
A system of grammars using symmetric phrase structure and translation rules in a Lisp version of Prolog is shown to provide symmetric bidirectional translation between English and Chinese for a fragment of the two languages.It is argued that symmetric grammars and translation rules significantly reduce the total grammar writing requirement for translation systems, and that research on symmetric translation systems deserves further study
C7820; C4210
computational linguistics;grammars;language translation;natural languages
translation;English;Chinese;symmetric phrase structure;Lisp version;Prolog;symmetric grammars

C87036513
Linguistic and extra-linguistic knowledge; a catalogue of language-related rules and their computational application in machine translation
SCHUBERT K.
Journal paper
Practical
ENG
ZZ
Comput.& Transl.(USA)
VOL. 1; NO. 3; PP. 125-52; 34 Ref.; DP. July-Sept. 1986
COMTE5
0884-0709
The author gives an overview of the problems encountered in translating a text.Language is a rule-governed system.Language science is the discovery, translation and application of these rules.But while a human translator can use the rules intuitively, the application of a computer involves the necessity of formulating the rules explicitly.Translation requires rules about both inside and outside influences on language;these rules in turn presume knowledge about those language related influences.After looking at the theoretical basis of this view he describes the practical details of the DLT machine translation system starting from the search for rules and knowledge.He sums up the rule systems and relates them to the types of knowledge they require.This concordance of rules and knowledge leads into a discussion of three characteristic features of the DLT system which might seem controversial, but which can then be shown to be strictly related to the rules and knowledge needed for machine translation.Finally the question of priority for either language-specific or extralinguistic rules and knowledge is taken up
C7820
computational linguistics;language translation;natural languages
extra linguistic knowledge;language related rules;machine translation;rules;knowledge

C87033349
Analysis method for natural language sentence
TSUJII J.
Kyoto Univ., Japan
Journal paper
Practical
JAP
JP
Inf.Process.Soc.Jpn.(Joho Shori) (Japan)
VOL. 27; NO. 8; PP. 924-32; 74 Ref.; DP. 1986
JOSHA4
0447-8053
To determine the 'structure' of an input sentence is the major problem of computational linguistics.When the 'structure' to be obtained finally is regarded as the structure required to represent the results of understanding, most of the studies on computational linguistics (including meaning processing and context processing) can be summarized into the analysis of a natural language sentence.The basic frames for sentence analysis, such as added context-free grammar, tree structure transformational grammar, etc.have been applied to practical systems, including mechanical translation systems.The author describes the current state and the basic problems concerning the studies of analysis methods for a natural language sentence
C4290; C7820
computational linguistics;grammars;language translation;natural languages
natural language sentence analysis;computational linguistics; understanding;meaning processing;context processing;added context free grammar;tree structure transformational grammar;mechanical translation systems

C87033343
A history and future prospects of computational linguistics
NAGAO M.
Kyoto Univ., Japan
Journal paper
General
JAP
JP
Inf.Process.Soc.Jpn.(Joho Shori) (Japan)
VOL. 27; NO. 8; PP. 855-61; 39 Ref.; DP. 1986
JOSHA4
0447-8053
Computational linguistics has been studied since the development of computers and has developed along with studies of mechanical translation.Mechanical translation was intensively studied from the latter half of 1950s but the US Congress concluded in 1965 that mechanical translation could not be materialized in a short period and that basic scientific studies on languages (computational linguistics) should be promoted (ALPAC Report).Computational linguistics includes the fields of acoustics, phonetics, phonology, morphology, lexicology, syntax, semantics, pragmatics, discourse, recognition and understanding, synthesis and generation, dialectology, translation, documentation writing aids, stylistics, content analysis, information retrieval, office automation, instruction, computer interfaces, graphics, speech, sign languages and animal languages
C4290; C7820
computational linguistics;history;language translation; technological forecasting
history;future prospects;computational linguistics;mechanical translation;acoustics;phonetics;phonology;morphology;lexicology; syntax;semantics;pragmatics;discourse;recognition;understanding, synthesis;generation;dialectology;documentation;writing aids; stylistics;content analysis;information retrieval;office automation;instruction;computer interfaces;graphics;speech;sign languages;animal languages

C87031189
Another stride towards knowledge-based machine translation
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
TOMITA M.; CARBONELL J. G.
Dept.of Comput.Sci., Carnegie-Mellon Univ., Pittsburgh, PA, USA
Conference paper
Practical
ENG
US
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 633-8; 28 Ref.; DP. 1986
Building on the well-established premise that reliable machine translation requires a significant degree of text comprehension, this paper presents a recent advance in multi-lingual knowledge-based machine translation (KBMT).Unlike previous approaches, the current method provides for separate syntactic and semantic knowledge sources that are integrated dynamically for parsing and generation.Such a separation enables the system to have syntactic grammars, language specific but domain general, and semantic knowledge bases, domain specific but language general.Subsequently, grammars and domain knowledge are precompiled automatically in any desired combination to produce very efficient and very thorough real-time parsers.A pilot implementation of the KBMT architecture using functional grammars and entity-oriented semantics demonstrates the feasibility of the new approach
C7820
computational linguistics;expert systems;grammars;language translation;natural languages
expert systems;knowledge based machine translation;text comprehension;semantic knowledge sources;parsing;generation; syntactic grammars;domain knowledge;functional grammars;entity oriented semantics

C87031187
Translation by understanding. a machine translation system LUTE
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany

25-29 Aug. 1986
NOMURA H.; NAITO S.; KATAGIRI Y.; SHIMAZU A.
NTT Basic Res.Labs., Tokyo, Japan
Conference paper
Practical
ENG
JP
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 621-6; 8 Ref.; DP. 1986
The article presents a linguistic model for language understanding and describes its application to an experimental machine translation system called LUTE.The language understanding model is an interactive model between the memory structure and a text.The memory structure is hierarchical and represented in a frame-network.Linguistic and non-linguistic knowledge is stored and the result of understanding the text is assimilated into the memory structure.The understanding process is interactive in that the text invokes knowledge and the understanding procedure interprets the text by using that knowledge.A linguistic model, called the extended case structure model, is defined by adopting three kinds of information: structure, relation and concept.These three are used recursively and iteratively as the basis for memory organization.These principles are applied to the design and implementation of the LUTE which translates Japanese into English and vice versa
C7820
computational linguistics;language translation;natural languages understanding;machine translation system;LUTE;linguistic model;memory structure;text;frame network;linguistic model;extended case structure model;structure;relation;concept;Japanese;English

C87031186
Semantic based generation of Japanese German translation system: result and evaluation
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
HANAKATA K.; LESNIEWSKI A.; YOKOYAMA S.
Inst.fur Inf., Stuttgart Univ., Germany
Conference paper
Practical
ENG
DE
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 560-2; 4 Ref.; DP. 1986
Project SEMSYN has achieved a state where a prototype system generates German texts on the basis of the semantic representation produced from Japanese texts by ATLAS/II of Fujitsu Laboratory.This paper describes some problems that are specific to the semantic based approach and some results of the evaluation study
C7820, C4290
computational linguistics;language translation
semantic based generation;Japanese German translation system;ATLAS II

C87031185
Controlled active procedures as a tool for linguistic engineering
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
LUCKHARDT H. D.; THIEL M.
Saarlandes Univ., Saarbrucken, Germany
Conference paper
Practical
ENG
DE
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 464-9; 20 Ref.; DP. 1986
Controlled active procedures are productions that are grouped under and activated by units called 'scouts'.Scouts are controlled by units called 'missions', which also select relevant sections from the data structure for rule application.Following the problem reduction method, the parsing problem is subdivided into ever smaller subproblems, each one of which is represented by a mission.The elementary problems are represented by scouts.The CAP grammar formalism is based on experience gained with natural language analysis and translation by computer in the Sonderforschungsbereich 100 at the University of Saarbrucken.The paper introduces CAP as a means of linguistic engineering
C7820; C4290
computational linguistics;language translation;natural languages controlled active procedures;rule writing;parsing strategies;linguistic engineering;relevant sections;data structure;rule application;problem reduction method;parsing problem;mission;scouts;CAP grammar formalism;natural language analysis;translation;Sonderforschungsbereich 100;University of Saarbrucken

C87029836
Model for lexical knowledge base
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
ISODA M.; AISO H.; KAMIBAYASHI N.; MATSUNAGA Y.
Keio Univ., Japan
Conference paper
Practical
ENG
JP
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 451-3; 7 Ref.; DP. 1986
Describes a model for a lexical knowledge base (LKB).An LKB is a knowledge base management system (KBMS) which stores various kinds of dictionary knowledge in a uniform framework and provides multiple viewpoints to the stored knowledge.KBMSs for natural language knowledge will be fundamental components of knowledgeable environments where non-computer professionals can use various kinds of support tools for document preparation or translation.However, basic models for such KBMSs have not been established yet.Thus, the authors propose a model for an LKB focusing on dictionary knowledge such as that obtained from machine-readable dictionaries
C6160; C7820
computational linguistics;database management systems;language translation;natural languages
lexical knowledge base;LKB;knowledge base management system;KBMS;dictionary knowledge;natural language knowledge;knowledgeable environments;support tools;document preparation;translation

C87031184
Acquisition of knowledge data by analyzing natural language
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
TANAKA Y.; YOSHIDA S.
Himeji Coll., Japan
Conference paper
Practical; Theoretical mathematical
ENG
JP
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 448-50; 3 Ref.; DP. 1986
Automatic identification of homonyms in kana-to-kanji conversion systems and of multivocal words in machine translation systems cannot be sufficiently implemented by the mere combination of grammar and word dictionaries.This calls for a new concept of knowledge data.What the new knowledge data is and how it can be acquired are mentioned in the paper.In natural language research, active discussion has been made within the framework of knowledge and samples of knowledge
C7820; C4290; C6130
computational linguistics;data acquisition;language translation;natural languages
knowledge data;natural language;homonyms;kana to kanji conversion systems;multivocal words;machine translation systems

C87031182
Experiments with an MT-directed lexical knowledge bank
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
PAPEGAAIJ B. C.; SADLER V.; WITKAM A. P. M.
BSO/Res.Bur.voor Systeemontwikkeling, Utrecht, Netherlands
Conference paper
Practical
ENG
NL
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 432-4; 6 Ref.; DP. 1986
A crucial test for any MT system is its power to solve lexical ambiguities.The size of the lexicon, its structural principles and the availability of extra-linguistic knowledge are the most important aspects in this respect.The paper outlines the experimental development of the SWESIL system: a structured lexicon-based word expert system designed to play a pivotal role in the process of distributed language translation (DLT) which is being developed in the Netherlands.It presents SWESIL's organizing principles, gives a short description of the present experimental set-up and shows how SWESIL is being tested at this moment
C7820
computational linguistics;expert systems;glossaries;language translation
MT directed lexical knowledge bank;MT system;SWESIL system;structured lexicon based word expert system;DLT;Netherlands

C87031181
Toward integrated dictionaries for M(a)T: motivations and linguistic organisation
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
BOITET C.; NEDOBEJKINE N.
GETA, Grenoble Univ., Saint-Martin-d'Heres, France
Conference paper
Practical
ENG
FR
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 423-8; 7 Ref.; DP. 1986
In the framework of machine aided translation systems, two types of lexical knowledge are used, 'natural' and 'formal', in the form of online terminological resources for human translators or revisors and of coded dictionaries for machine translation proper.A new organization is presented, which allows both types to be integrated in a unique structure, called 'fork' integrated dictionary, or FID.A given FID is associated with one natural language and may give access to translations into several other languages.The FIDs associated to languages L1 and L2 contain all information necessary to generate coded dictionaries of M(a)T systems translating from L1 into L2 or vice-versa

C7820
computational linguistics;glossaries;language translation
fork integrated dictionary,lexical database;specialized language, linguistic programming;integrated dictionaries;linguistic organisation;machine aided translation systems;lexical knowledge; online terminological resources;human translators;coded dictionaries;FID;natural language,M a T systems

C87030068
CRITAC- a Japanese text proofreading system
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
TAKEDA K.; FUJISAKI T.; SUZUKI E.
Japan Sci.Inst., IBM Japan Ltd., Tokyo, Japan
Conference paper
Practical
ENG
JP
Inst,Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 412-17; 19 Ref.; DP. 1986
CRITAC (critiquing using accumulated knowledge) is an experimental expert system for proofreading Japanese text.It detects mistypes, Kana-to-Kanji misconversions, and stylistic errors.This system combines Prolog-coded heuristic knowledge with conventional Japanese text processing techniques which involve heavy computation and access to large language databases
C7106; C7820
computational linguistics;expert systems;language translation;word processing
Japanese text proofreading system;CRITAC;critiquing using accumulated knowledge,experimental expert system;mistypes;Kana to Kanji misconversions;stylistic errors;Prolog coded heuristic knowledge;conventional Japanese text processing techniques; language databases

C87029702
SCSL: a linguistic specification language for MT
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
ZAJAC R.
GETA, Grenoble Univ., Saint-Martin-d'Heres, France
Conference paper
Practical
ENG
FR
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 393-8; 10 Ref.; DP. 1986
Nowadays, MT systems grow to such a size that a first specification step is necessary if one wants to be able to master their development and maintenance, for the software part as well for the linguistic part ('lingwares').Advocating for a clean separation between linguistic tasks and programming tasks, the paper introduces a specification/implementation/validation framework for NLP then SCSL, a language for the specification of analysis and generation modules
C6140D; C7820
computational linguistics;language translation,natural languages, specification languages
machine translation;natural language processing;specification language;linguistic specification language;MT systems;linguistic tasks;programming tasks;specification implementation validation framework;NLP;SCSL,generation modules

C87029701
PeriPhrase: lingware for parsing and structural transfer
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
BEESLEY K. R.; HEFNER D.
A,L,P,Syst., Provo, UT, USA
Conference paper
Practical
ENG
US
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 390-2; 0 Ref.; DP. 1986
PeriPhrase is a high-level computer language developed by A.L.P.Systems to facilitate parsing and structural transfer.It is designed to speed the development of computer-assisted translation systems and grammar checkers.The syntax and semantics of this tool are described together with its integrated development environment
C6140D; C7820
computational linguistics,high level languages,language translation
lingware,parsing,structural transfer,high level computer language, computer assisted translation systems;grammar checkers;syntax; semantics;integrated development environment

C87028459
Pragmatics in machine translation
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
ROTHKEGEL A.
Saarbrucken Univ., Germany
Conference paper
Theoretical mathematical
ENG
DE
Inst.Angewandte Kommunikations & Sprachforschung,Bonn, Germany
NP. xix+675; PP. 335-7; 12 Ref.; DP. 1986
TEXAN is a system of transfer-oriented text analysis.Its linguistic concept is based on a communicative approach within the framework of speech act theory.In this view texts are considered to be the result of linguistic actions.It is assumed that they control the selection of translation equivalents.The transition of this concept of linguistic actions (text acts) to the model of computer analysis is performed by a context-free elocution grammar processing categories of actions and a propositional structure of states of affairs.The grammar which is related to a text lexicon provides the connection of these categories and the linguistic surface units of a single language
C4290; C7820
computational linguistics;language translation
machine translation;TEXAN;transfer oriented text analysis; linguistic concept;communicative approach;speech act theory; translation equivalents;linguistic actions;text acts;computer analysis;context free elocution-grammar;states of affairs; linguistic surface units

C87031178
Strategies for interactive machine translation. the experience and implications of the UMIST Japanese project
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
WHITELOCK P. J.; MCGEE WOOD M.; CHANDLER B. J.; HOLDEN N.; HORSFALL H. J.
Centre for Comput.Linguistics, Univ.of Manchester Inst.of Sci.& Technol., England
Conference paper
General; Practical
ENG
GB
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 329-34; 27 Ref.; DP. 1986
The authors are designing an English-to-Japanese interactive machine translation system.For development purposes we are using an existing corpus of 10000 words of continuous prose from the ICL PERQ's graphics documentation;in the long term, the system will be extended for use by technical writers in fields other than software.The authors have developed system development software, user interface, and grammar and dictionary handling utilities.The English analysis grammar handles most of the syntactic structures of the corpus, and there are a range of formats for output of linguistic representations and Japanese text.A transfer grammar for English-Japanese has been prototyped, but is not yet fully adequate to handle all constructions in the corpus;a facility for dictionary entry in kanji is incorporated.The authors focus on its interactive nature, discussing the range of different types of interaction which are provided or permitted for different types of user
C7820
computational linguistics,language translation;natural languages; user interfaces
English to Japanese translation;interactive machine translation; UMIST Japanese project;graphics documentation;user interface, grammar;dictionary handling utilities;English analysis grammar, syntactic structures;linguistic representations;Japanese text;kanji

C87031177
NARA: a two-way simultaneous interpretation system between Korean and Japanese-a methodological study
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
HEE SUNG CHUNG; KUNII T. L.
Dept.of Inf.Sci., Tokyo Univ., Japan
Conference paper
Practical; Theoretical mathematical; Experimental
ENG
JP
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 325-8; 7 Ref.; DP. 1986
The authors present a new computing model for constructing a two-way simultaneous interpretation system between Korean and Japanese.They also propose several methodological approaches to the construction of a two-way simultaneous interpretation system, and realize the two-way interpreting process as a model unifying both linguistic competence and linguistic performance.The model is verified theoretically and through actual applications
C7320; C4290
computational linguistics;language translation;natural languages
NARA,two way simultaneous interpretation system,Korean,Japanese, two way interpreting,process;linguistic competence,linguistic performance

C87031176
Idioms in the Rosetta machine translation system
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
SCHENK A.
Philips Res.Labs., Eindhoven, Netherlands
Conference paper
Theoretical mathematical
ENG

NL
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 319-24; 12 Ref.; DP. 1986
Rosetta is an experimental translation system which uses an intermediate language and translates between Dutch, English and, in the future, Spanish.The theoretical framework of Rosetta which is based on isomorphic M-grammars is outlined.Idioms are then discussed in this framework.Finally some examples are considered
C7820; C4290
computational linguistics;language translation;natural languages
Rosetta machine translation system;Dutch;English;Spanish;isomorphic M grammars

C87031175
A compositional approach to the translation of temporal expressions in the Rosetta system
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
APPELO L.
Philips Res.Labs., Eindhoven, Netherlands
Conference paper
Theoretical mathematical
ENG
NL
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 313-18; 12 Ref.; DP. 1986
This paper discusses the translation of temporal expressions, in the framework of the machine translation system Rosetta.The translation method of Rosetta, the 'isomorphic grammar method', is based on Montague's Compositionality Principle.It is shown that a compositional approach leads to a transparent account of the complex aspects of time in natural language and can be used for the translation of temporal expressions
C7820; C4290
computational linguistics;language translation;natural languages
compositional approach;temporal expressions.Rosetta system;machine translation system;isomorphic grammar method;Compositionality Principle;natural language

C87031174
Valency theory in a stratificational MT-system (machine translation)
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
SCHMIDT P.
IAI Eurotra-D, Saarbrucken, Germany
Conference paper
Theoretical mathematical
ENG
DE
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 307-12; 7 Ref.; DP. 1986
The author investigates valency theory as a linguistic tool in machine translation.There are three main areas in which major questions arise.(1) Valency theory itself.He sketches a valency theory in linguistic terms which includes the discussion of the nature of dependency representation as an interface for semantic description.(2) The dependency representation in the translation process.He sketches the different roles of dependency representation in analysis and generation.(3) The implementation of valency theory in an MT-system.He gives a few examples for how a valency description could be implemented in the Eurotra-formalism
C7820; C4290
computational linguistics;language translation;valency
valency theory;linguistic tool;machine translation;dependency representation;semantic description;valency description;Eurotra formalism

C87031173
Generating semantic structures in Eurotra-D
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
STEINER E.
IAI-Eurotra-D, Saarbrucken, Germany
Conference paper
Theoretical mathematical
ENG
DE
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 304-6; 14 Ref.; DP. 1986
Analysis and generation of clauses within the Eurotra-framework proceeds through the levels of (at least) Eurotra constituent structure (ECS), Eurotra relational structure (ERS) and interface structure (IS).At IS, labelling of nodes consists of labellings for time, modality, semantic features, semantic relations and others.In this paper, we shall be concerned exclusively with semantic relations (SRs) or participant roles (PR).In Eurotra-D, they have been experimenting with a set of SRs, or PRs, which are identified with the help of syntactic criteria.This approach is outlined
C7820; C4290
computational linguistics;language translation
clause generation;semantic structures;constituent structure;relational structure;interface structure;semantic relations;participant roles;syntactic criteria

C87028458
The (C,A), T framework in Eurotra: a theoretically committed notation for MT (machine translation)
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
ARNOLD D. J.; KRAUWER S.; ROSNER M.; DES TOMBE L.; VARILE G. B.
Essex Univ., Colchester, England
Conference paper
Theoretical mathematical
ENG
GB
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 297-303; 10 Ref.; DP. 1986
This paper describes a model for MT, developed within the Eurotra MT project, based on the idea of compositional translation, by describing a basic, experimental notation which embodies the idea.Some of the theoretical and practical implications of the model, including some concrete extensions, and some more speculative aspects are discussed
C4290; C7820
computational linguistics;language translation;research initiatives
Eurotra MT project;compositional translation

C87031172
Linguistic developments in Eurotra since 1983
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
JASPAERT L.
Katholieke Univ.Leuven, Belgium
Conference paper
General
ENG
BE
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; PP. 294-6; 6 Ref.; DP. 1986
The author puts the theory and metatheory currently adopted in the Eurotra project into a historical perspective, indicating where and why changes to its basic design for a transfer-based machine translation (TBMT) system have been made
C7820; C4290
computational linguistics;language translation;research initiatives
linguistics;monostratal;dimensionality;Eurotra project;transfer based machine translation

C87028449
11th International Conference on Computational Linguistics.Proceedings of Coling '86
11th International Conference on Computational Linguistics.Proceedings of Coling '86
Bonn, Germany
25-29 Aug. 1986
Conference proceedings
Practical; Theoretical mathematical
ENG
ZZ
Inst.Angewandte Kommunikations & Sprachforschung;Bonn, Germany
NP. xix+675; DP. 1986
The following topics were dealt with: sign-theoretical model of semantic structure;computational analysis;linguistic semantics;temporal relations;term associations in automatic information retrieval;lexical data;PeriPhrase, lingware for parsing and structural transfer;SCSL, linguistic specification language for MT;ATN programming environment;CRITAC, Japanese text proofreading system;integer codes for text storage;BetaText, event-driven text processing and text analyzing system;integrated dictionaries for M(a)T;word database for national language processing;automatic thesaurus construction;and functional structures for parsing dependency constraints.Abstracts of individual papers can be found under the relevant classification codes in this or other issues
C4290; C7820
computational linguistics;language translation;natural languages
sign theoretical model;semantic structure;computational analysis;linguistic semantics;temporal relations;term associations;automatic information retrieval;lexical data;PeriPhrase;lingware;parsing;structural transfer;SCSL;linguistic specification language;MT;ATN programming environment;CRITAC;Japanese text proofreading system;integer codes;text storage;BetaText;event driven text processing;text analyzing system;integrated dictionaries;M a T;word database;national language processing;automatic thesaurus construction;functional structures;dependency constraints

C87031148
An information model for a machine translation system
LEONT EVA N. N.
Journal paper
Practical; Theoretical mathematical
ENG
ZZ
Nauchno-Tekh.Inf.Ser.2 (USSR)
Autom.Doc.& Math.Linguist.(USA)
VOL. 19; NO. 10; PP. 22-9; 5 Ref.; DP. 1985
VOL. 19; NO. 5; PP. 92-105; DP. 1985
NIPSBP
ADMLAE
0548-0027
0005-1055
0005-1055/85/$20.00
The French-Russian machine translation system developed at the

All-Union Translation Center is based on an information model for understanding any natural-language text.The paper deals with the external and internal reasons for choosing this model, together with the composition of it, which involves a discussion of certain key problems in the linguistic support.It is shown that the strategy of combining information and translation functions at first sight complicates the task, whereas in fact it relieves some load on the system and makes the task feasible, while flexible links between the components mean that one can adapt the system to difference topic areas and various information requirements on the text
C7820; C4290
computational linguistics;language translation;natural languages
natural language text understanding·information model;French Russian machine translation system;All Union Translation Center; linguistic support;information requirements

C87028444
Semantic modules in a machine translation system: complex term analysis
BELYAEVA L. N.; MATORINA L. V.; PIOTROVSKII R. G.; YASHCHENKO T. V.
Journal paper
Practical; Theoretical mathematical
ENG
ZZ
Nauchno-Tekh.Inf.Ser.2 (USSR)
Autom.Doc.& Math.Linguist.(USA)
VOL. 19; NO. 4; PP. 29-34; 8 Ref.; DP. 1985
VOL. 19; NO. 4; PP. 52-61; DP. 1985
NIPSBP
ADMLAE
0548-0027
0005-1055
0005-1055/85/$20.00
A basic task handled in the machine translation MT of scientific texts is to extract the basic meaning from the input text with minimal loss and transmit it correctly by means of output-language facilities.The solution is largely determined by the correct identification of compound terms (noun terminological work combinations) in the text, followed by analysis and translation.This is so because these word combinations contain the main professional information in the text and reflect the basic scientific concepts in the area of knowledge represented by it.The speech-statistics group has designed and implemented four approaches to the machine translation of noun word combinations, which have been operated routinely and on an experimental basis: translating noun terminological combinations as a whole without analysis by reference to a phrasal dictionary;translation on the basis of a component analysis for each wordform in the noun word combination;matrix-free translation;thesaurus-network translation
C4290; C7820
computational linguistics;language translation
semantic modules;semiotics;engineering linguistics;computational linguistics;machine translation;complex term analysis;scientific texts;compound terms;noun terminological work combinations;speech statistics;noun word combinations;phrasal dictionary;component analysis;matrix free translation;thesaurus network translation

C87026400
Discussion session on machine translation
Alvey/ICL Workshop on Linguistic Theory and Computer Applications.Transcripts of Presentation and Discussions (CCL/UMIST-86/2)
Manchester, England
Sept. 1985
JOHNSON R.; WHITELOCK P.(Ed.); SOMERS H.(Ed.); BENNETT P.(Ed.); JOHNSON R.(Ed.)
WOOD M. M.(Ed.)
Centre for Computational Linguistics, Univ.of Manchester Inst.of Sci.Tecnnol., England
Conference paper
Practical
ENG
GB
Univ.Manchester Inst.Sci.& Technol;Manchester, England
NP. 215; PP. 169-89; 8 Ref.; DP. March 1986
MT raises some quite interesting theoretical methodological questions which haven't really been raised up to now.The author concentrates on that particular collection of issues.The author assumes that MT systems are extensive, that translation is from natural language, they apply to more than two languages, they don't require human intervention and they are linguistic based
C7820
computational linguistics;language translation;natural languages
machine translation;MT systems;natural language

C87026398
Alvey/ICL Workshop on Linguistic Theory and Computer Applications.Transcripts of Presentation and Discussions (CCL/UMIST-86/2)
Alvey/ICL Workshop on Linguistic Theory and Computer Applications.Transcripts of Presentation and Discussions (CCL/UMIST-86/2)
Manchester, England
Sept. 1985
WHITELOCK P.(Ed.); SOMERS H.(Ed.); BENNETT P.(Ed.); JOHNSON R.(Ed.); WOOD M. M.(Ed.)
Conference proceedings
Practical
ENG
ZZ
Univ.Manchester Inst.Sci.& Technol;Manchester, England
NP. 215; DP. March 1986
The following topics were dealt with: linguistic analysis and linguistic theory;default inheritance;deterministic parsing;grammars;machine translation;lexicons;and syntax and semantics.Abstracts of individual papers can be found under the relevant classification codes in this or other issues
C7820; C4210
computational linguistics;grammars;language translation;natural languages
natural languages;Alvey;linguistic analysis,linguistic theory; default inheritance;deterministic parsing;grammars,machine translation;lexicons;syntax;semantics

C87013908
Computer **aided** **translation** project, University Sains Malaysia, Penang, Malaysia
WAROTAMASIKKHADIT U.
Journal paper
General
ENG
ZZ
Comput.& Transl.(USA)
VOL. 1; NO. 2; PP. 113; 0 Ref.; DP. April-June 1986
COMTES
0884-0709
Research in CAT started in 1978 with development of grammar models for English-Malay translation using the software tool ARIANE.A basic translation system with a vocabulary of 1000 lexical units was completed in 1982.In 1984, a permanent CAT project unit was established, and a laboratory prototype for English-Malay translation was successfully tested in 1985.The English-Thai machine translation project in Thailand was established in June 1981.Two committees have been appointed to undertake this task, the English-Thai translation research project, and the Thai structures research project for English-Thai machine translation using the ARIANE system
C7820
language translation
English Malay translation;software tool ARIANE;English Thai machine translation project

C87004517
Linguistic research in the Greek group (for Eurotra project)
TSITSOPOULOS S.
Journal paper
General; Practical
ENG
ZZ
Multilingua (Netherlands)
VOL. 5; NO. 3; PP. 149-51; 6 Ref.; DP. 1986
MULTDF
0167-8507
0167-8507/86/$2.00
Work for Eurotra in Greece began in a double vacuum, the lack of a substantial body of theoretical work on the Greek language inspired by contemporary linguistic paradigms, and the total absence of ongoing programmes, academic or otherwise, in any branch of computational linguistics.These infrastructural deficiencies, normally distinct and with independent histories, converge disconcertingly in an MT project
C7820
language translation;linguistics;natural languages
linguistic research;Greek group;Eurotra project;Greek language. linguistic paradigms;computational linguistics;MT project

C87004511
Eurotra: general overview
PERSCHKE S.
CEC, Luxembourg
Journal paper
General
ENG
LU
Multilingua (Netherlands)
VOL. 5; NO. 3; PP. 134-5; 0 Ref.; DP. 1986
MULTDF
0167-8507
0167-8507/86/$2.00
Eurotra is a multilingual machine translation project carried out by the Commission of the European Communities.The article shows not only the intrinsic scientific interest and ambition of the Eurotra project, but also its impact on the future of computational linguistics in Europe
C7820
language translation;linguistics
multilingual machine translation project;Commission of European Communities;Eurotra project;computational linguistics

C86057816
Sentence disambiguation by asking
TOMITA M.
Dept.of Comput.Sci., Carnegie-Mellon Univ., Pittsburgh, [illegible] USA
Journal paper
Practical; Theoretical mathematical
ENG
US
Comput.& Transl.(USA)
VOL. 1; NO. 1; PP. 39-51; 9 Ref.; DP. Jan.-March 1986
0884-0709
Describes a technique for asking questions to disambiguate a sentence.Such a disambiguation technique is crucial for interactive machine translation systems, and helps resolve structural ambiguities.The shared-packed forest representation and

the forest shaving algorithm, along with the efficient parsing algorithm, enable us to parse and disambiguate highly ambiguous sentence with hundreds of parses without dealing with hundreds of individual parse trees
C7820; C4210; C4290
computational linguistics;grammars;language translation
sentence disambiguation,disambiguation technique;interactive machine translation systems;structural ambiguities;shared packed forest representation;forest shaving algorithm;efficient parsing algorithm;parse trees

C86057814
Language, sublanguage, and the promise of machine translation
BARON N. S.
Southwestern Univ., Georgetown, TX, USA
Journal paper
Practical; Theoretical mathematical
ENG
US
Comput.& Transl.(USA)
VOL. 1; NO. 1; PP. 3-19; 9 Ref.; DP. Jan -March 1986
0884-0709
Looks at machine translation (and at natural language processing more generally) in context of a model of linguistic communication.In developing this model, the author discovered strong parallels between human-human communication on the one hand and human-machine (or human-human communication mediated by machines, as in the case of machine translation) on the other
C7820; C4290; C4210
computational linguistics;formal languages;language translation; natural languages
human machine communication;machine translation;natural language processing;linguistic communication;human human communication

B86054473; C86043853
Review of text-to-speech conversion technology
SAGISAKA Y.; SATO H.
NTT Res.Labs., Musashino, Japan
Journal paper
Practical
JAP
JP
J.Acoust.Soc.Jpn.(Japan)
VOL. 41; NO. 12; PP. 901-5; 42 Ref ; DP. Dec. 1985
NIOGAH
0369-4232
This paper is concerned with text-to-speech conversion technology, including text analysis and speech synthesis.For text analysis necessary for speech synthesis, high-grade semantic analysis is needed for **computer**-**aided** **translation**.However, it is necessary to improve the semantic analysis technology to some extent for efficient application
B6130; C5585
speech synthesis
text to speech conversion technology;speech synthesis;computer aided translation;semantic analysis

C86033676
Augmented dependency grammar: a simple interface between the grammar rule and the knowledge
Second Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Geneva, Switzerland
27-29 March 1985
MURAKI K.; ICHIYAMA S.; FUKUMOCHI Y.
C&C Systems Res.Lab., NEC Corp., Kawasaki, Japan
Assoc.Comput.Linguistics
Conference paper
Practical; Theoretical mathematical
ENG
JP
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. vii+276; PP. 198-204; 2 Ref.; DP. 1985
This paper describes some operational aspects of as language comprehension model which unifies the linguistic theory and the semantic theory with respect to operations.The computational model, called augmented dependency grammar (ADG), formulates not only the linguistic dependency structure of sentences but also the semantic dependency structure using tha extended deep case grammar and field-oriented fact-knowledge based interferences.Fact knowledge base and ADG model clarify the qualitative difference between what we call semantics and logical meaning.From a practical view point, it provides clear image of syntactic/semantic computation for language processing in analysis and synthesis.It also explains the gap in semantics and logical meaning, and gives a clear computational image of what is called conceptual analysis
C4290; C4210; C7820
computational linguistics;grammars;language translation
augmented dependency grammar;interface;grammar rule;knowledge; operational aspects;language comprehension model;linguistic theory; semantic theory;computational model;ADG;semantic dependency structure;language processing

C86036381
An English generator for a case-labelled dependency representation
Second Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Geneva, Switzerland
27-29 March 1985
TAIT J. I.
Acorn Computers Ltd., Cambridge, England
Assoc.Comput.Linguistics
Conference paper
Practical
ENG
GB
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. vii+276; PP. 194-7; 12 Ref.; DP. 1985
The paper describes a program which has been constructed to produce English strings from a case-labelled dependency representation.The program uses an especially simple and uniform control structure with a well defined separation of the different knowledge sources used during generation.Furthermore, the majority of the system's knowledge is expressed in a declarative form, so in principle the generator's knowledge bases could be used for purposes other than generation.The generator uses a two-pass control structure, the first translating from the semantically orientated case labelled dependency structures into surface syntactic trees and the second translating from these trees into English strings
C7820; C4290
computational linguistics;language translation
English generator;case labelled dependency representation;program; control structure;knowledge bases;two pass control structure; surface syntactic trees

C86036379
A probabilistic parser
Second Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Geneva, Switzerland
27-29 March 1985
GARSIDE R.; LEECH F.
Lancaster Univ., England
Assoc.Comput.Linguistics
Conference paper
Practical; Theoretical mathematical
ENG
GB
Assoc.Comput.Linguistics,Morristown, NJ, USA
NP. vii+276; PP. 165-70; 6 Ref.; DP. 1985
The UCREL team at the University of Lancaster is engaged in the development of a robust parsing mechanism, which will assign the appropriate grammatical structure to sentences in unconstrained English text.The techniques used involve the calculation of probabilities for competing structures, and are based on the techniques successfully used in tagging (i.e.assigning grammatical word classes) to the LOB (Lancaster-Oslo/Bergen) corpus.The first step in the parsing process involves dictionary lookup of successive pairs of grammatically tagged words, to give a number of possible continuations to the current parse.Since this lookup will often not be able unambiguously to distinguish the point at which a grammatical constituent should be closed, the second step of the parsing process will have to insert closures and distinguish between alternative parses.It will generate trees representing these possible alternatives, insert closure points for the constituents, and compute a probability for each parse tree from the probability of each constituent within the tree
C7820; C4290
computational linguistics;grammars;language translation
LOB corpus;probabilistic parser;UCREL,University of Lancaster; robust parsing mechanism;grammatical structure;unconstrained English text,dictionary lookup;trees;closure points

C86036378
A probabilistic approach to grammatical analysis of written English by computer
Second Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Geneva, Switzerland
27-29 March 1985
BEALE A. D.
Bowland Coll., Lancaster Univ., England
Assoc.Comput.Linguistics
Conference paper
Practical
ENG
GB
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. vii+276; PP. 159-65; 5 Ref.; DP. 1985
Work at the Unit for Computer Research on the English Language at the University of Lancaster has been directed towards producing a grammatically annotated version of the Lancaster-Oslo/Bergen (LOB) Corpus of written British English texts as the preliminary stage in developing computer programs and data files for providing a grammatical analysis of unrestricted English text.Work is now in progress to devise a suite of programs to provide a constituent analysis of the sentences in the corpus.So far, sample sentences have been automatically assigned phrase and clause tags using a probabilistic system similar to word tagging.It is hoped that the entire corpus will eventually be parsed
C7820; C4280
computational linguistics;language translation
Lancaster Oslo Bergen corpus;LOB corpus;probabilistic approach; grammatical analysis;written English;University of Lancaster; computer programs;data files;phrase;clause

C86036377
'Lexifanis' a lexical analyzer of modern Greek
Second Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Geneva, Switzerland
27-29 March 1985
KOTSANIS Y.; MAISTROS Y.

Dept.of Comput.Sci., Nat.Tech.Univ.of Athens, Greece
Assoc.Comput.Linguistics
Conference paper
Practical; Theoretical mathematical
ENG
GR
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. vii+276; PP. 154-8; 12 Ref.; DP. 1985
Lexifanis is a software tool designed and implemented by the authors to analyze modern Greek language.This system assigns grammatical classes (parts of speech) to 95-98% of the words of a text which is read and normalized by the computer.By providing the system with the appropriate grammatical knowledge (i.e.: dictionaries of non-inflected words, affixation morphology and limited surface syntax rules) any 'variant' of modern Greek language (dialect or idiom) can be processed
C7820; C4290; C6115
computational linguistics;language translation;software tools
lexical analyzer;modern Greek;software tool;grammatical classes, dictionaries;affixation morphology,limited surface syntax rules

C86035060
Design and implementation of a lexical data base
Second Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Geneva, Switzerland
27-29 March 1985
WEHRLI E.
Dept.of Linguistics, California Univ., Los Angeles, CA, USA
Assoc.Comput.Linguistics
Conference paper
Theoretical mathematical
ENG
US
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. vii+276; PP. 146-53; 17 Ref.; DP. 1985
This paper is concerned with the specifications and the implementation of a particular concept of word-based lexicon to be used for large natural language processing systems such as machine translation systems, and compares it with the morpheme-based conception of the lexicon traditionally assumed in computational linguistics.It is argued that, although less concise, a relational word-based lexicon is superior to a morpheme-based lexicon from a theoretical, computational and also practical viewpoint
C61600; C4290, C4250
computational linguistics;language translation;natural languages; relational databases
lexical data base;specifications;implementation;word based lexicon; natural language processing systems;machine translation systems; morpheme based conception;computational linguistics;relational word based lexicon

C86036373
A two-way approach to structural transfer in MT
Second Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Geneva, Switzerland
27-29 March 1985
ROOT R.
Linguistics Res.Center, Texas Univ., Austin, TX, USA
Assoc.Comput.Linguistics
Conference paper
Practical; Theoretical mathematical
ENG
US
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP vii+276; PP. 70-2. 4 Ref.; DP. 1985
The METAL machine translation project incorporates two methods of structural transfer-direct transfer and transfer by grammar.The author discusses the strengths and weaknesses of these two approaches in general and with respect to the METAL project, and argues that, for many applications, a combination of the two is preferable to either alone
C7820; C4290
computational linguistics;language translation
two way approach;structural transfer;METAL machine translation project;direct transfer;transfer by grammar

C86036372
An evaluation of METAL: the LRC machine translation system
Second Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Geneva, Switzerland
27-29 March 1985
SLOCUM J.; BENNETT W. S.; WHIFFIN L.; NORCROSS E.
Assoc.Comput.Linguistics
Conference paper
Practical
ENG
ZZ
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. vii+276; PP. 62-9; 2 Ref.; DP. 1985
The Linguistics Research Center (LRC) at the University of Texas at Austin is currently developing METAL, a fully-automatic high quality machine-translation system.This paper describes the current status of METAL, emphasizing the results of the most recent post-editors' evaluation, and briefly indicates some future directions for the system.A six-page German original text and a raw (unedited, but automatically reformatted) METAL translation of that text into English are included as appendices
C7820; C4290
computational linguistics;language translation
METAL;fully automatic high quality machine translation system

C86033667
Second Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Second Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Geneva, Switzerland
27-29 March 1985
Assoc.Comput.Linguistics
Conference proceedings
ENG
ZZ
Second Conference of the European Chapter of the Association for Computational Linguistics.Proceedings of the Conference
Assoc.Comput.Linguistics;Morristown, NJ, USA
NP. vii+276; DP. 1985
The following topics were dealt with: natural languages;grammars;machine translation;parsing;Boolean operators;lexical database;probabilistic parsar;computational theory;database queries;automated speech recognition;sentence production model;user modelling;dialog structure, and dialog strategy in Ham-Ans;communicative context, of dialogue interaction structure.Abstracts of individual papers can be found under the relevant classification codes in this or other issues
C4290; C4210; C7820
computational linguistics;grammars;language translation,natural languages
natural languages;grammars;machine translation;parsing;Boolean operators;lexical database;probabilistic parsar;computational theory;database queries;automated speech recognition;sentence production model;user modelling;dialog structure;dialog strategy; Ham Ans;communicative context;dialogue interaction structure

C86036369
Expert system for semantic disambiguation of ill-formed text in an Esperanto-based intermediate language for machine translation (SWESIL)
FANTOM I. D.
Polytech.South Bank, London, England
Dissertation
Practical
ENG
GB
NP. 148
The distributed language translation project employs a modified subset of Esperanto as an intermediate language (IL) for machine translation of information text between natural languages.Text entered in the source language (SL) is analysed syntactically by an SL module and then passed to an IL module for semantic disambiguation.The task of the semantic module is to identify the most plausible syntactic parse.Interleaved (online) semantics, in which the syntactic and semantic modules have a symbiotic relationship, is employed.Word meanings are represented by semantic vectors, and plausibility is expressed using Zaden's test-score semantics and fuzzy logic techniques.Design principles are developed on a basis of the literature on psycholinguistics, semantics and computational linguistics and an expert system design, in Prolog, is presented.A further aspect of the work is an expert system for consistency control, using vectorial cross-checking techniques.A simplified subset of the system is implemented in MicroProlog, and some preliminary results on the disambiguation of twenty-four meanings of 'Time flies' (without syntactic pre-processing) provids grounds for encouragement in the further development of the system
C7820; C4290
computational linguistics;distributed processing;expert systems, language translation
distributed language translation project;Esperanto;machine translation;information text;natural languages;source language, semantic disambiguation.syntactic parse;symbiotic relationship, semantic vectors,Zaden s test score semantics;fuzzy logic;expert system;consistency control;vectorial cross checking

C86026472
A survey of machine translation: its history, current status, and future prospects
SLOCUM J.
Microelectron.& Comput.Technol.Corp., Austin, TX, USA
Journal paper
Bibliography; General; Practical
ENG
US
Comput.Linguist.(USA)
VOL. 11; NO. 1; PP. 1-17, 56 Ref.; DP. Jan.-March 1985
CLINEE
0362-613X
0362-613X/85/010001-17$03.00
Elements of the history, state of the art, and probable future of machine translation (MT) are discussed.The treatment is largely tutorial.The paper covers some of the major MT R&D groups, the general techniques they employ, and the roles they play in the development of the field.The conclusions concern the seeming permanence of the translation problem, and potential re-integration of MT with mainstream Computational Linguistics
C7820
language translation;linguistics;technological forecasting
future prospects;history;machine translation;MT;R D groups; Computational Linguistics

C86026488
The construction theory and practice of a syntax analyzer used in the English-Chinese translation experimental system
WANG KAIZHU; FENG YIN
Journal paper

Practical; Theoretical mathematical
CHI
ZZ
J.Harbin Inst.Technol.(China)
NO. 3; PP. 8-15; 5 Ref.; DP. 1985
HPKYAY

The syntax analyzer mechanism possesses two distinguishing characteristic features: the grammars of the two languages and the programmes are separated strictly, and a presage mechanism and an exceeding search mechanism are used instead of the technology of 'backup' in parsing.Theoretically, the construction process of these two mechanisms is described, an algorithm of building this presage mechanism-a presage analysis table is given, and a series of techniques to improve the presage ability in the grammar reformation are introduced.In order to prove the feasibility of the design, the authors have built a syntax analyzer used in the English-Chinese translation system, and obtained important verification data

C7820; C4290
computational linguistics;language translation
construction theory;syntax analyzer;English Chinese translation experimental system;presage mechanism;exceeding search mechanism; algorithm

C86020903
Interaction of syntactic and semantic structures during linguistic analysis
KUDRYASHOVA I. M.; SOKOLOVA E. G.
Journal paper
Practical
ENG
ZZ
Nauchno-Tekh.Inf.Ser.2 (USSR)
Autom.Doc.& Math.Linguist.(USA)
VOL. 18; NO. 6; PP. 17-24 ; 18 Ref.; DP. 1984
VOL. 18; NO. 3; PP. 77-89; DP. 1984
NIPSBP
ADMLAE
0548-0027
0005-1055
0005-1055/85/$20.00

Describes the organization of the interaction between syntactic and semantic structures in a system for automatic translation from French to Russian (FRAT) developed at the All-Union Translation Center (AUTS).The FRAT language apparatus is based on two metalanguages used for linguistic analysis: a syntactic language describing the form of the utterance, and a semantic language explicating the content.Information about the sentence is extracted gradually: first the syntactic analysis system constructs the primary syntactic representation of the sentence without using semantic information to do so;then the semantic analysis system constructs the primary semantic representation, interpreting the primary syntactic representation of the sentence in terms of the semantic metalanguage using syntactic-semantic dictionaries.The primary representations complement each other, and the syntactic metalanguage is used for exchange of information.The article describes the initial stage of the interaction of these two representations of the sentence

C7820; C4290
computational linguistics;language translation;linguistics
syntactic structures;information exchange,semantic structures; linguistic analysis;automatic translation;French;Russian;FRAT;All Union Translation Center;AUTS;FRAT language apparatus; metalanguages;syntactic language;semantic language;sentence: syntactic analysis system;primary syntactic representation: semantic analysis system;primary semantic representation;syntactic semantic dictionaries

C86015834
Relevance, points of view and dialogue modelling
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
WILKS Y.; NIRENBURG S.(Ed.)
Comput.Res.Lab., New Mexico State Univ., Las Cruces, NM, USA
Conference paper
Practical
ENG
US
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 370-87; 27 Ref.; DP. 1985

This paper compares two approaches to the modelling of human discourse and, more particularly, dialogue.Both place themselves within a general information processing paradigm, and both descend for the insights of Grice (1975) that understanding is a matter of inference from what is said and what is assumed.So general is that assumption now, and so widespread are the disciplines that draw upon it (philosophy, psychology, linguistics and artificial intelligence (AI)) that it is hard to capture briefly except in opposition to the transformational-generative paradigm of language, with its notions of the primacy and autonomy of syntax, and the theoretical primacy of explications of competence over those of performance

C7820; C4290
computational linguistics;language translation;natural languages
relevance;points of view;dialogue modelling;human discourse; dialogue;information processing paradigm;understanding;inference; philosophy;psychology;linguistics;artificial intelligence;syntax

C86015833
Characteristics of the METAL machine translation system at production stage
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
WHITE J. S.; NIRENBURG S.(Ed.)
Conference paper
Practical
ENG
ZZ
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 359-69; 12 Ref.; DP. 1985

The METAL machine translation system, a joint project of the Linguistic Research Center and Siemens, has been released for use as part of marketed translation systems.The system, which presently translates technical German into English, is an outgrowth of a traditional, generative approach to automatic analysis and synthesis of natural language phenomena carried on at the Linguistics Research Center for many years.In its present manifestation, it is a modular design consisting of purely monolingual lexicons, transfer lexicons, and an augmented phrase structure grammar The grammar is powerful enough to constrain application, to build new nodes with essential characteristics of their sons and new synthetic information as well, and to perform transformations to re-order, delete, and create constituents.The parsar is enhanced to allow application of rules in levels, and eliminating unlikely paths via preferential weightings calculated from lexical and grammatical data.The METAL system, conceived in recent years as destined for implementation, has an orientation to user interface which includes sophisticated text stripping, unfound word handling and reconstitution, and a convenient means of working with the lexicons interactively

C7820; C4290; C4210
computational linguistics;grammars:language translation;natural languages
MT;METAL;machine translation system;Linguistic Research Center: Siemens:technical German;English;natural language;monolingual lexicons;transfer lexicons;augmented phrase structure grammar; parsar;preferential weightings;grammatical data;user interface: text stripping;unfound word handling

C86015832
Reflections on the knowledge needed ' process ill-formed language
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
WEISCHEDEL R. M.; RAMSHAW L. A.; NIRENBURG S.(Ed.)
Bolt, Beranek, & Newman Inc., Cambridge, MA, USA
Conference paper
Practical
ENG
US
Colgate Univ Hamilton, NY, USA
NP. 387; PP 348-58; 14 Ref.; DP 1985

This paper reflects about the kinds of morphological, syntactic. semantic, and pragmatic knowledge needed to process ill-formed input The authors conclude that an excellent start on processing ill-formed input has been exemplified in a number of concrete implementations, but that a substantial amount of fundamental work must still be done if systems are to understand language robustly to the degree that humans do.Furthermore, they conclude that studying ill-formed language offers important perspectives on the knowledge and architecture needed to correctly understand natural languages

C7820; C4290
computational linguistics.language translation;natural languages
morphological language;syntactic knowledge;semantic knowledge:MT; knowledge;ill formed language;pragmatic knowledge;understand language;natural languages

C86014958
Knowledge resource tools for accessing large text files
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
WALKER D. E.; NIRENBURG S.(Ed.)
Artificial Intelligence & Inf.Sci,Res., Bell Commun.Res., Morristown, NJ, USA
Conference paper
Practical
ENG
US
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 335-47; 20 Ref.; DP. 1985

This paper provides an overview of a research program defined at Bellcore.The objective is to develop facilities for working with large document collections that provide more refined access to the information contained in these 'source' materials than is possible through current information retrieval procedures.The tools being used for this purpose are machine-readable dictionaries, encyclopedias, and related 'resources' that provide geographical, biographical, and other kinds of specialized knowledge.A major feature of the research program is the exploitation of the reciprocal relationships between sources and resources.These interactions between texts and tools are intended to support experts who organize and use information in a workstation environment.Two systems under development are described to illustrate the approach: one providing capabilities for full-text subject assessment;the other for concept elaboration while reading text.Progress in the research depends critically on developments in artificial intelligence, computational linguistics, and information science to provide a scientific base, and on software engineering, database management, and distributed systems to

provide the technology
C7250; C4290; C6160; C7820
computational linguistics;database management systems;expert systems;information retrieval;language translation;natural languages
MT;large text files;Bellcore;large document collections; information retrieval;machine readable dictionaries;encyclopedias; sources;resources;experts;workstation environment;full text subject assessment;reading text;artificial intelligence; computational linguistics;information science;software engineering; database management;distributed systems

C86015831
On debugging environment proposed for EUROTRA
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
VERASTEGUI N.; NIRENBURG S.(Ed.)
Inst.de Formation et Conseil en Inf., Grenoble, France
Conference paper
Practical
ENG
FR
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 323-34; 5 Ref.; DP. 1985
A proposal of external specification of the user environment for the EUROTRA project is presented.The needs of the users and the functions which are necessary for any efficient testing environment are analyzed
C7820; C4290
computational linguistics;language translation;natural languages; user interfaces
MT;interfaces;debugging environment;EUROTRA;external specification; user environment;users;testing environment

C86013089
Static grammars, a formalism for the description of linguistic models
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
VAUQUOIS B.; CHAPPUY S.; NIRENBURG S.(Ed.)
Conference paper
Practical
ENG
ZZ
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 298-322; 0 Ref.; DP. 1985
For a linguistic model it is necessary, first of all, to define the mapping between the strings of words of a language and their structural organisation, given that with transducers there are many ways of obtaining the same result using different strategies.This mapping called a static grammar is independent of the analysis, generation or whatever strategy adopted.Moreover the formalism of a static grammar is not affected by the choice or number of interpretation levels.The authors present a static grammar formalism.Using this formalism, any given language can be described as a series of 'charts'.Each 'chart' describes how a certain group of strings corresponds to the structure associated with this group of strings (this structure is a valid and complete substructure of the linguistic model).The structures of all the sentences of a language for a given linguistic model can be described by means of a series of chart inter-references.The static grammar is used as a base for writing dynamic analyses and generation modules, however, the static grammar does not concern itself with strategic, combinatorial, ambiguity problems or the choice of structures related to dynamic grammars
C4290; C7820; C4210
computational linguistics;grammars;language translation;natural languages
MT;formalism;linguistic models;linguistic model;transducers;static grammar;chart inter references;dynamic analyses;generation modules; dynamic grammars

C86015830
Feasibility study of personal/interactive machine translation systems
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
TOMITA M.; NIRENBURG S.(Ed.)
Dept.of Comput.Sci., Carnegie-Mellon Univ., Pittsburgh, PA, USA
Conference paper
Practical
ENG
US
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 289-97; 11 Ref.; DP. 1985
Most existing practical machine translation systems are designed to translate documentation, such as technical papers and manuals.However, there is a growing need for translating not only large texts but also personal short texts such as letters and informal messages.The conventional machine translation systems, which are intended to translate large texts, are not very suitable for these kinds of small jobs.One needs an interactive system which has a totally different design philosophy.This paper describes the design philosophy of personal/interactive machine translation system, and studies its feasibility
C7820
computational linguistics;interactive systems;language translation; natural languages;personal computing;user interfaces
personal interactive machine translation systems;documentation; technical papers;manuals;personal short texts;letters;informal messages;interactive system

C86015829
A preliminary linguistic framework for EUROTRA, June 1985
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
DES TOMBE L.; ARNOLD D.; JASPAERT L.; JOHNSON R.; KRAUWER S. ROSNER M.; VARILE N.; WARWICK S.; NIRENBURG S.(Ed.)
Conference paper
Practical
ENG
ZZ
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 283-8; 0 Ref.; DP. 1985
The work described here was the consequence of the idea that the authors wanted to make a new, more interesting theoretical start in EUROTRA.It is preliminary and not fully developed yet;it should be seen as the reflection of a way of thinking about MT.Currently, they are making it more precise, and experimenting with it.They sketch the general outlines of the new EUROTRA framework
C7820; C4290
computational linguistics;language translation;natural languages
linguistic framework;EUROTRA;MT

C86015828
Linguistics and natural language processing
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
RASKIN V.; NIRENBURG S.(Ed.)
Purdue Univ., Lafayette, IN, USA
Conference paper
Practical
ENG
US
Colgate Univ.,Hamilton, NY, USA
NP. 387; PP. 268-82; 8 Ref.; DP. 1985
The paper addresses the issue of cooperation between linguistics and natural language processing (NLP), in general, and between linguistics and machine translation (MT), in particular.It focuses on just one direction of such cooperation, namely applications of linguistics to NLP, virtually ignoring for now any possible applications of NLP to linguistics, which can range from providing computer-based research tools and aids to linguistics to implementing formal linguistic theories and verifying linguistic models.The author deals with the question why linguistics must be applied to NLP and what the consequences of ignoring it are.He provides a counterpoint by discussing how linguistics should not be applied to NLP and, by contrast and inference, how it should be.He narrows the discussion down to one promising approach to NLP, the sublanguage deal, and the interesting ways, in which linguistics can be utilized within a limited sublanguage.He discusses the things linguistics can contribute to MT
C7820; C4290
computational linguistics;language translation;natural languages
natural language processing;linguistics;machine translation, computer based research tools;formal linguistic theories.verifying linguistic models;sublanguage

C86015827
The level hypothesis in discourse analysis
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation.of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
PUSTEJOVSKY J.; NIRENBURG S.(Ed.)
Dept.of Comput.& Inf.Sci., Massachusetts Univ., Amherst, MA, USA
Conference paper
Practical
ENG
US
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 245-67; 21 Ref.; DP. 1985
The author explores some difficult questions related to topics in discourse analysis (henceforth DA) and offers a partial solution to some of them.In particular, he addresses the issue of levels in DA and how the various approaches taken within the field can be classified according to a leveled model.He then considers an approach for representing the semantics of discourse, and considers how it fits in to the proposed model for DA
C7820; C4290
computational linguistics;language translation;natural languages
MT;linguistics;natural languages;level hypothesis;discourse analysis;semantics

C86015826
Interlingua design for TRANSLATOR
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
NIRENBURG S.; RASKIN V.; TUCKER A. B.; NIRENBURG S.(Ed.)
Dept.of Comput.Sci., Colgate Univ., Hamilton, NY, USA
Conference paper
Practical
ENG

US
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 224-44; 17 Ref.; DP. 1985
The interlingua approach to machine translation (MT) is characterized by the following two stages: (1) translation of the source text into an intermediate representation, an artificial language (interlingua) which is designed to capture the various types of meaning of the source text and (2) translation from the interlingua into the target text.Over the years a number of MT projects tried to develop interlingua-based systems.In these projects the amount of linguistic and encyclopaedic knowledge used to produce intermediate representations was quite limited However, even at that level difficulties connected with encoding knowledge seemed overwhelming.The TRANSLATOR project at Colgate University benefits from recent developments in knowledge representation techniques.The text of its interlingua text reflects syntactic, lexical, contextual, discourse (including speech situation) and pragmatic meaning of the input.This paper discusses the lexicon and grammar of the interlingua used in TRANSLATOR, and touches upon the structure of the bilingual (source language to interlingua) dictionaries
C7820; C4290; C4210
computational linguistics;expert systems;grammars;language translation;natural languages
TRANSLATOR;interlingua approach;machine translation;artificial language;meaning;source text;target text;encyclopaedic knowledge; Colgate University;knowledge representation techniques;discourse; pragmatic meaning;lexicon;grammar;dictionaries

C86015825
Structural transformation in the generation stage of MU Japanese to English machine translation system
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
NAGAO M.; NIRENBURG S.(Ed.)
Dept.of Electr.Eng., Kyoto Univ., Japan
Conference paper
Practical
ENG
JP
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 200-23; 4 Ref.; DP. 1985
The author outlines the Mu MT project, looking at the general principles and the linguistic framework.He then discusses the transfer from Japanese to English, looking at dependency structure, target language word selection, global sentential structures and inference and context
C7820
computational linguistics;language translation;natural languages parsing;grammars;MU Japanese;English machine translation system; linguistic framework;dependency structure;target language word selection;global sentential structures;inference;context

C86015824
Recovering the speaker's decisions during mechanical translation
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
MCDONALD D. D.; NIRENBURG S.(Ed.)
Massachusetts Univ., Amherst, MA, USA
Conference paper
Practical
ENG
US
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 183-99; 9 Ref.; DP. 1985
When studied as a source of insight into the human language faculty, rather than to construct a commercially useful service, mechanical translation (MT) is carried out by coupling an otherwise normal natural language parsing system to a normal natural language generation system.The author proposes that a crucial capability has been omitted from the design of the parsers that have been used to date, namely a facility for recognizing the information that is implicit in the form of any well written text;matters of emphasis, whether a fact is new or old, whether a relationship is given explicitly or left as an obvious inference, signals of intended moves in the discourse, and other things of this sort.He claims that mechanical translations are 'mechanical' principally because they pay no attention to information of this sort, and propose that this can be dealt with by incorporating into the parser knowledge of the relationship between usage and form of the sort that is commonplace in any modern language generation system
C7820; C4290; C4210
computational linguistics;grammars;language translation;natural languages
speaker s decisions;mechanical translation;human language faculty; natural language parsing system;natural language generation system; well written text;emphasis;inference;modern language generation system

C86015823
LMT: a PROLOG-based machine translation system
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
MCCORD M. C.; NIRENBURG S.(Ed.)
IBM Thomas J.Watson Res.Center, Yorktown Heights, NY, USA
Conference paper
Practical
ENG
US
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 179-82; 3 Ref ; DP. 1985
The author describes a machine translation system, LMT, based in PROLOG, translating from English to German.The effort on LMT per se has just begun this year, although the logic programming methodology for the analysis of the source (English) goes back several years
C7820; C4290; C6150C; C4210
computational linguistics;grammars;language translation,natural languages;program compilers
compilers;grammars;linguistics;MT;top down parsing;LMT;PROLOG based machine translation system;English;German;logic programming

C86015822
Integrating syntax and semantics
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
LYTINEN S. L.; NIRENBURG S.(Ed.)
Cognitive Syst.Inc., New Haven, CT, USA
Conference paper
Practical
ENG
US
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 167-78; 12 Ref.; DP. 1985
Well-known examples such as Bar-Hillel's (1960) 'The box is in the pen' illustrate that extensive semantic analysis is necessary to resolve ambiguities that must be resolved in machine translation.If one accepts the premise that semantics should be added to the analysis techniques used in machine translation, what is the way in which it should be added?This paper argues for an integrated approach to semantic processing.That is, syntactic and semantic processing should take place at the same time, rather than in separate stages.However, although the author argues for the integration of syntactic and semantic analysis processes, he also argues for the use of a separate body of syntactic knowledge, and for building a separate syntactic representation during the parsing process.This is in contrast to previous integrated parsers, which have relied almost exclusively on semantic representations to guide the parsing process, and which have not used a separate body of syntactic rules
C7820; C4290; C4210
computational linguistics,grammars;language translation;natural languages
linguistics;natural language;syntax;semantics,machine translation; analysis techniques;integrated approach;semantic processing; syntactic knowledge;syntactic representation;parsing process

C86015821
The significance of sublanguage for automatic translation
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
KITTREDGE R. I.; NIRENBURG S.(Ed.)
Montreal Univ., Que. Canada
Conference paper
Practical
ENG
CA
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 154-66; 13 Ref.; DP. 1985
This paper addresses three questions: what is sublanguage?;why is sublanguage analysis important for automatic translation?;and how can a translation system take advantage of sublanguage properties?The first of these questions appears to have a simple answer.Natural languages clearly have specialized varieties which are used in reference to restricted subject matter.One speaks, for example, of the 'language of chemistry' to mean a loosely defined set of sentences or texts dealing with a particular part of reality.But when one considers the automatic translation of specialized language, one is forced to be more precise.One must describe sublanguages as coherent, rule-based systems.The attempt to write grammars for special-purpose sublanguages raises a number of theoretical and practical problems, which are only now being discussed.But since the only path to high-quality automatic translation seems to lie through sublanguage (at least during the next decade or two), one has no choice but to solve these problems
C7820; C4290; C4210
computational linguistics,grammars;language translation;natural languages
sublanguage;automatic translation;rule based systems;grammars

C86015820
Machine translation as an expert task
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
JOHNSON R.; WHITELOCK P.; NIRENBURG S.(Ed.)
Centre for Comput.Linguistics, Univ.of Manchester Inst.of Sci.& Technol., England
Conference paper
Practical.
ENG
GB
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 145-53; 8 Ref.; DP. 1985

The case against fully automatic high quality machine translation (FAHQMT) has been well-canvassed in the literature ever since ALPAC.Although considerable progress in computational linguistics has been made since then, many of the major arguments against FAHQMT still hold.Accepting that FAHQMT is not possible in the current state of the art, it is both feasible and desirable to set up R&D programmes in MT which can both produce results which will satisfy sponsors and provide an environment to support research directed towards bringing MT closer to the ultimate goal of FAHQMT
C7820; C4290
computational linguistics;expert systems,language translation, natural languages
natural languages;ES;expert task;high quality machine translation; computational linguistics;MT

C86015819
Machine translation in the SDCG formalism
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
XIUMING HUANG; NIRENBURG S.(Ed.)
Inst.of Linguistics, Chinese Acad.of Social Sci., Beijing, China
Conference paper
Practical
ENG
CN
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 135-44; 10 Ref.; DP. 1985
The paper describes the SDCG (semantic definite clause grammars), a formalism for natural language processing (NLP), and the XTRA (English Chinese sentence translator) machine translation (MT) system based on it.The system translates general domain English sentences into grammatical Chinese sentences in a fully automatic manner.It is written in Prolog and implemented on the DEC-10, the GEC, and the SUN workstation.SDCG is an augmentation of DCG (definite clause grammars) which in turn is based on CFG (context free grammars).Implemented in Prolog, the SDCG is highly suitable for NLP in general, and MT in particular.A wide range of linguistic phenomena is covered by the XTRA system, including multiple work senses, coordinate constructions, the prepositional phrase attachment, among others
C7820; C4290, C4210
computational linguistics,grammars;language translation;natural languages
SDCG formalism;SDCG;semantic definite clause grammars,natural language processing;XTRA,English Chinese sentence translator, machine translation;English sentences;grammatical Chinese sentences;Prolog;DEC 10,GEC,SUN workstation;definite clause grammars;context free grammars;linguistic phenomena;multiple work senses;coordinate constructions,prepositional phrase attachment

C86015818
On the design of expert systems grafted on MT systems
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
GERBER R.; BOITET C.; NIRENBURG S.(Ed.)
Genoble Univ., Saint-Martin-d'Heres, France
Conference paper
Practical
ENG
FR
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 116-34; 17 Ref.; DP. 1985
MT systems integrate many advanced concepts from the fields of computer science, linguistics, and AI; specialized languages for linguistic programming based on production systems, complete linguistic programming environment, multilevel representations, organization of the lexicons around 'lexical units', units of translation of the size of several paragraphs, possibility to use text-driven heuristic strategies.The authors are now beginning to integrate new techniques: unified design of an 'integrated' lexical data-base containing the lexicon in 'natural' and 'coded' form, use of the 'static grammars' formalism as a specification language, and design of a kind of structural metaeditor (driven by some static grammar) allowing the interactive construction of a document.This paper centers on the study on possible additions of expert systems equipped with metalinguistic and extralinguistic knowledge, in order to solve some problems encountered in second-generation MT systems.Several examples of the possible use of expert-corrector systems in M(a)T (machine (aided) translation) systems are given
C7820
computational linguistics;expert systems;language translation; natural languages
expert systems;MT systems,computer science,linguistics;AI, linguistic programming;production systems,programming environment, multilevel representations;lexicons;translation;text driven heuristic strategies;lexical data base;grammars;specification language;structural metaeditor;expert systems;extralinguistic knowledge;expert corrector systems

C86015817
Lexicon-driven machine translation
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
CULLINGFORD R. E.; ONYSHKEVYCH B. A.; NIRENBURG S.(Ed.)
Georgia Inst.of Technol., Atlanta, GA, USA
Conference paper
Practical
ENG
US
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 75-115; 24 Ref.; DP. 1985
Machines translation (MT) systems have historically relied upon explicit grammars in order to analyze the source text and reproduce it in the target language.The authors argue for a style of MT in which the focus of processing is at the level of the lexicon, rather than the grammar.This approach to translation allows an analyzer to map source sentences into an interlingual form, which then can be mapped (perhaps after intermediate inferencing steps) back into target sentence(s) which are paraphrase-equivalent to the original.Advantages of the approach include: (1) the possibility for different paraphrases of the original;(2) the capability for multi-sentence expression of the original when no single work (e.g.a verb) exists in the target language which spans the same meaning complex as a word in the source;(3) a uniform approach to word sense disambiguation and anaphoric reference resolution;and, most importantly, (4) the possibility for robust handling of ungrammatical and ellipsed source text
C7820; C4290
computational linguistics,language translation,natural languages
linguistics;natural language;machine translation;explicit grammars; target language;lexicon;source sentences;inferencing steps; paraphrases;word sense disambiguation;anaphoric reference resolution;ellipsed source text

C86015816
New approached to machine translation
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
CARBONELL J. G.; TOMITA M.; NIRENBURG S.(Ed.)
Carnegie-Mellon Univ., Pittsburgh, PA, USA
Conference paper
Practical
ENG
US
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 59-74; 21 Ref.; DP. 1985
The current resurgence of interest in machine translation is partially attributable to the emergence of a variety of new paradigms, ranging from better translation aids and improved pre- and post-editing methods, to highly interactive approaches and fully automated knowledge-based systems.This paper discusses each basic approach and provides some comparative analysis.It is argued that both interactive and knowledge based systems offer considerable promise to remedy the deficiencies of the earlier, more ad-hoc post-editing approaches
C7820
computational linguistics;expert systems;language translation; natural languages
machine translation;translation aids;post editing methods, automated knowledge based systems

C86015815
A case study in software evolution. from ARIANE-78.4 to ARIANE-85
Proceedings of the Conference on Theoretical and-Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
BOITET C.; GUILLAUME P.; QUEZEL AMBRUNAZ M.; NIRENBURG S.(Ed.)
Grenoble Univ., Saint-Martin-d'Heres, France
Conference paper
Practical
ENG
FR
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 27-58; 16 Ref.; DP. 1985
ARIANE-78 has been used for years at GETA as the underlying programming environment for writing many MT systems for subsystems, in a set of specialized (rule based) languages for linguistic programming (SLLP).The authors present its recent evolution, which has been prompted by the feedback from the users, and has led the implementors to a deep reshaping.In particular, the control structure of the entire environment has been parametrized to a large extent, due to the introduction of a specialized (finite state based) language used for describing sets of possible sequences of linguistic processes (phases'), such as structural analysis or lexical expansion
C7820; C4290; C6115
computational linguistics,language translation;natural languages; programming environments
machine translation;rule based languages finite state languages; software evolution,ARIANE 78 4,ARIANE 85,programming environment, MT systems,linguistic programming,control structure,linguistic processes;structural analysis;lexical expansion

C86015814
On the production environment proposed for the EUROTRA project
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
BACHUT D.; NIRENBURG S.(Ed.)
Groupe d'Etudes pour la Traduction Auto.-Univ.Sci.et Med.de Grenoble, Saint Martin d'Heres, France
Conference paper
Practical
ENG

FR
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 15-26; 8 Ref.; DP. 1985
The author presents the general architecture of a production environment which is specific for a M(A)T system, and gives some proposals to integrate new functionalities in this system.A good management of the results of the translation process may lead to an easier improvement of the linguistic data.he describes a possible organisation for the machine environment of such a system for the management of the data base of texts.Finally, he gives some general rules for the implementation of a monitor
C7820; C4290; C6160Z
computational linguistics;database management systems;language translation;natural languages
text database;production environment;EUROTRA project;translation process;linguistic data;machine environment;monitor

C86015813
A Mul view of the <C,A>, T framework in EUROTRA
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
ARNOLD D.; JASPAERT L.; JOHNSON R.; KRAUWER S.; ROSNER M.
DES TOMBE L.; VARILE N.; WARWICK S.; NIRENBURG S.(Ed.)
Conference paper
Practical
ENG
ZZ
Colgate Univ.;Hamilton, NY, USA
NP. 387; PP. 1-14; 4 Ref.; DP. 1985
The background to this paper is the attempt within EUROTRA to develop a general framework for research and development work in MT, providing in particular an environment which facilitates reasoning about the relationships between the representations that are necessary for automatic translation between natural languages.The more immediate background is the attempt to apply this framework experimentally on a small scale in developing a 'proto-EUROTRA'.This paper gives a reasonably clear idea about the user language and theories of representation for this experiment, and to indicate en route some of the directions for further work.It reports work in progress, and is thus deliberately speculative, programmatic, and rather informal
C7820; C4290
computational linguistics;language translation;natural languages
linguistics;EUROTRA;MT;automatic translation,natural languages, user language

C86015812
Proceedings of the Conference on Theoretical and Methocological Issues in Machine Translation of Natural Languages
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Hamilton, NY, USA
14-16 Aug. 1985
NIRENBURG S.(Ed.)
Conference proceedings
Practical
ENG
ZZ
Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages
Colgate Univ.;Hamilton, NY, USA
NP. 387; DP. 1985
The following topics were dealt with: EUROTRA;ARIANE;expert systems;lexicon driven MT;grammars;sublanguage;inference and context;TRANSLATOR;natural language processing,linguistics,and METAL.Abstracts of individual papers can be found under the relevant classification codes in this or other issues
C7820; C4290
computational linguistics;language translation;natural languages
machine translation;EUROTRA;ARIANE;expert systems;lexicon driven MT,grammars;sublanguage;inference;context;TRANSLATOR;natural language processing;linguistics;METAL

C86015810
Automatic generation and analysis of Chinese in machine translation
Proceedings of the 1984 South East Asia Regional Computer Conference.SEARCC 84
Hong Kong
24-28 Sept. 1984
ZHIWEI F
Inst.of Sci.& Technol., Beijing, China
Conference paper
General; Theoretical mathematical
ENG
CN
SEARCC;Hong Kong
NP. 493; PP. 19/1-25; ? Ref.; DP 1984
The automatic generation and analysis of Chinese is the central topic of machine translation in China.This paper describes the intermediate constituent method and the logico-semantic method for the generation of Chinese, it describes also the multi-label and multi-branch tree method for the analysis of Chinese.At the same time, it deals with some special problems in Chinese such as the generation of measure words and the reorganization of the word order of multi-modifiers
C7820; C1250
computational linguistics;computerised pattern recognition; language translation
Chinese;machine translation;China;intermediate constituent method; logico semantic method;multi branch tree method;measure words;word order

C86011049
Computer-**aided** **translation** at WCC
PERSCHEID M. M.
Journal paper
Practical
ENG
ZZ
CALICO J.(USA)
VOL. 3; NO. 1; PP. 22-4; 0 Ref.; DP. Sept. 1985
CALJE8
0742-7778
Many individuals, companies, and government agencies need to have a large volume of foreign language printed matter translated for their use.**Computer**-**aided** **translation** offers the advantages of speed and volume over the normal non-assisted human translation process.Weidner is one company which recognized the needs in this particular area and has developed both hardware and software to fill this need.Constant improvement and attention to detail is needed to keep such a system operating at top accuracy It is Weidner's goal to stay on top of all advances in the field as well as to offer a complete line of language and translation services to the community
C7820
language translation
computer aided translation;WCC;foreign language printed matter; Weidner;translation services

C86011069
Computational linguistics projects at the Centre for Computational Linguistics
Journal paper
Practical
ENG
ZZ
Multilingua (Netherlands)
VOL. 4; NO. 3, PP 170-1; 0 Ref , DP. 1985
MULTDF
0167-8507
0167-8507/85/0004-0170$2.00
Some computational linguistics projects at CCL are outlined.They include English-Japanese machine translation, MT systems, and work on EUROTRA
C7820; C4290
computational linguistics;language translation;linguistics
Centre for Computational Linguistics,computational linguistics projects;CCL;English Japanese machine translation;MT systems; EUROTRA

C86011068
Computational linguistics projects at the University of Saarbrucken
Journal paper
Practical
ENG
ZZ
Multilingua (Netherlands)
VOL. 4; NO. 3; PP. 165-71; 29 Ref.; DP. 1985
MULTDF
0167-8507
0167-8507/85/0004-0165$2.00
Computational linguistics projects at Saarbrucken University are outlined.These include the SUSY translation system, the ASCOF system, and the TEXAN text analysis system
C7820; C4290
computational linguistics;educational courses,language translation; linguistics;text editing
computational linguistics;Saarbrucken University;SUSY translation system;TEXAN text analysis system

C86041522
Machine **translation** in Europe and America
NOMURA H.
NTT Res.Lab., Japan
Journal paper
General
JAP
JP
Inf.Process.Soc.Jpn.(Joho Shori) (Japan)
VOL. 26; NO. 10; PP. 1226-36; 3 Ref.; DP. 1985
JOSHA4
0447-8053
Principal movements for the research and development on machine **translation** in Europe and America are described based on reports from 1982-3.The terminology data bank of the Commission of the European Communities (CEC), machine **translation**, the EUROTRA project, machine **translation** project of the French Government (ADI TAO-ESOPE), a machine **translation** experimental system ARIANE-78, a machine **translation** system called **TITUS**, a terminology data bank called LEXIS and a terminology data bank TERMIUM are outlined
C7820
language translation
machine translation,terminology;CEC;EUROTRA;ADI TAO ESOPE;ARIANE 78;TITUS;terminology data bank;LEXIS;TERMIUM

C86020908
TITUS IV.System for automatic and simultaneous **translation** in four languages
DUCROT J. M.
Journal paper
Practical
GER
ZZ

Sprache & Datenverarb.(Germany)
VOL. 9; NO. 1; PP. 28-36; 0 Ref.; DP. 1985
SPDADH
0343-5202
TITUS IV is a machine-**translation** system which has evolved from the needs of multilanguage documentation.The texts are formulated as abstracts from documents and can be rendered in German, English, French or Spanish.They are simultaneously translated into the other languages.In order to minimize the difficulties which result from the complexity of natural languages, the system is based upon a controlled syntax.The **translation** ensues exclusively in dialogue.Any uncorrected clauses require interaction with the editor.The elements of the dictionaries and related syntactical structures are described in detail
C7820
language translation;linguistics
simultaneous translation;TITUS IV;machine translation systen; multilanguage documentation;controlled syntax;syntactical structures

C86004413
Interactive editing of technical documents with a view to automatic **translation**
IDT 85.Information, Documentation, Transfert des Connaissances.6 Congres National sur l'Information et la Documentation (IDT '85.Information, Documentation, Transfer of Knowledge.6th National Congress on Information and Documentation)
Versailles, France
12-14 June 1985
ZAJAC R.
CNRS, Saint-Martin-D'Heres, France
Conference paper
Practical
FRE
FR
Assoc.Francaise des Documentalistes et des Bibliothecaires Specialises;Paris, France
NP; 279; PP. 18-24; 11 Ref.; DP 1985
2-900195-05-5
A method aimed at eventual full automation of machine **translation** of technical documents is discussed, wherein the editor uses a computer to sort out and correct textual ambiguities.In particular three approaches (ITS, **TITUS**, and EPISTLE) are considered in some detail
C61300; C7820
interactive systems;language translation;text editing
interactive editing;technical documents;machine translation; textual ambiguities;ITS;TITUS;EPISTLE

C85047913
New developments in **TITUS** 4
Tools for the Trade : Translating and the Computer 5.Proceedings of a conference
London, England
10-11 Nov. 1983
STREIFF A. A.; LAWSON V.(Ed.)
Dept.of Inf.& Data Process., Inst.Textile de France, Boulogne-Billancourt, France
Aslib;Translators' Guild;Comm.Eur Communities
Conference paper
Practical
ENG
FR
Aslib;London, England
NP. xi+272; PP. 185-92; 0 Ref.; DP 1985
0-85142-180-6
The **TITUS** 4 system was originally designed to produce abstracts in the form of sentences or phrases written in controlled syntax.It is now being improved, partly to give the user more flexibility in writing sentences, and partly so that the system can be implemented in other fields than abstracting services.Improvements being introduced to enhance **TITUS** 4's versatility include multiple-clause sentences.Certain restrictions, however, remain owing to linguistic problems associated with **translation** from one languago to another
C7820; C7240
abstracting;language translation
TITUS 4;abstracts;controlled syntax,multiple clause sentences. linguistic problems;translation

C85021610
Machine **translation** and peripherals, evaluation, post-editing, attitudes and training
LOFFLER LAURIAN A. M.
CNRS, Paris, France
Journal paper
General; Practical
FRE
FR
Contrastes (France)
spec; ser.A4; PP. 43-67; 16 Ref.; DP. Jan. 1984
CNTRDQ
0247-815X
The author explains that machine **translation** (MT) is not yet acceptable without post-editing which is examined in some depth:How much should the post-editor intervene, who is he and what training has he had?In a brief look at Systran and **Titus** she mentions evaluation and quality assessment of translations, the EURODICAUTOM data base and the pros and cons of language pair systems as opposed to multilingual
C7820; C7130
government data processing;language translation
language pair systems;multilingual systems;peripherals;evaluation; post editing;attitudes;training;machine translation;MT;Systran; Titus;quality assessment;translations;EURODICAUTOM data base

C84041252
Computer assisted **translation** (TAO) at the Centre de Documentation Scientifique et Technique (CDST) of the Centre National de la Recherche Scientifique (CNRS)
DETEMPLE A.
Journal paper
Practical
ENG
ZZ
Multilingua (Netherlands)
VOL. 2; NO. 4; PP. 189-94; 3 Ref.; DP. 1983
MULTDF
0167-8507
0167-8507/83/0002-0189$2.00
Putting the PASCAL documentary [illegible] onto a multilingual footing poses a difficult problem.Several computer **translation** systems likely to provide suitable solutions have been tested in recent years at the Centre de Documentation Scientifique et Technique (CDST) of the French Centre National de la Recherche Scientifique.There are two systems already operational that might be able to produce translations suitable for CDST purposes.**TITUS** IV, devised by the Institut Textile de France, is semiconversational and has certain input restrictions, so that it takes about 30 minutes to input a 50 to 60-word summary.The American SYSTRAN system, of which the EEC has acquired the rights for certain European language pairs, is fully computerized.However, it requires a certain amount of postediting and the staff of the CDST is currently attempting to establish exactly how much Among the other systems studied, ALPS, marketed in France by Control Data under the name of TRANSMATIC, involves conversational processes designed to deal with all types of language processing.The second generation tools, GETA and SYGMART offer the most possibilities
C7820; C7430; C6110
language translation;program and system documentation
CDST;CNRS;PASCAL documentary base;computer translation systems; TITUS IV;American SYSTRAN system;postediting;ALPS;Control Data. TRANSMATIC;GETA;SYGMART

C83043941
TITUS IV: a system for the automatic and simultaneous **translation** of four languages
Information Management Research in Europe.Proceedings of the EURIMS Conference
Versailles, France
May 1982
DUCROT J M.; TAYLOR P. J (Ed.); CRONIN B.(Ed.)
Inst.Textile de France, Boulogne-Billancourt, France
Conference paper
Application; Practical
FRE
FR
Aslib;London, England
NP. 212; PP. 177-95; 0 Ref.; DP. 1983
0-85142-171-7
TITUS IV is an automatic **translation** technique intended to manipulate scientific and technical articles with terms in German, English, French and Spanish.Basic concepts of controlled areas representing the vocabularies and the syntactic rules are described.The procedure consists of feeding each input sentence with a language code to a multilingual lexicon, forming a pivot language, indexing, transforming the grammar and producing 2, 3 or 4 translated texts.The vocabulary includes subgroups of specialist terms.The pivot language is in binary form.The lexicon caters for 4 main grammatical forms, namely substantives, adjectives, verbs and adverbs, although there are several other groups.Error messages are printed in all four languages, enabling the operator to redraft the input as required.Applications of the technique have been made using an IBM 4311-12, the major processing of each document occupying only 2.5 secs.of CPU time.At the terminal, however, a typical input and output schedule would be 25 documents, including bibliographic references, during 6 hours work
C7820
language translation
simultaneous translation;languages;automatic translation,German. English;French;Spanish,vocabularies;syntactic rules,input sentence. multilingual lexicon;pivot language;indexing;grammar-substantives. adjectives;verbs;adverbs

C80002470
Information system **TITUS**
GURTLER Z.
Suoro, Praha, Czechoslovakia
Journal paper
General
CZE
CS
Cesk.Inf.Teor.& Praxe (Czechoslovakia)
VOL. 21; NO. 6; PP. 175-81; 4 Ref.; DP. 1979
CITPBH
General characteristics of the system are given together with the description of the variant **TITUS** II which is presently used.The variant based on a formalized documentary language can be processed by the computer and makes the automated **translation** of processed document records stored in the computer memory into German, English, Spanish and French, possible
C7210; C7820
information services;language translation
TITUS;documentary language;automated translation;German;English; Spanish;French

C79031242
The development of the '**TITUS**' four-language automatic **translation** method
STEIFF S.
Journal paper
Application; Practical
FRE
ZZ
Inf.& Doc.(France)
NO. 4; PP. 20-6; 0 Ref.; DP. May 1979
CDIDAQ
'**TITUS** II' is essentially based on a documentary language which is a simplified and formalised form of natural language.The author describes the vocabulary principles used and the standard structure of phrases.The mode of operation is schematically illustrated and the role and the output of translators is described
C7820
language translation
TITUS;automatic translation method;documentary language;natural language;four language translation

---

C78032034
Experiences with **TITUS** II
ZINGEL H. J.
ZTDI, Dusseldorf, Germany
Journal paper
General; Practical
ENG
DE
Int.Classif.(Germany)
VOL. 5; NO. 1; PP. 33-7; 0 Ref.; DP March 1978
Description of the international cooperative documentation system called **TITUS** (Textile Information Treatment Users' Service) in its previous and present form (**TITUS** II).It uses a special linguistic way of automatic **translation** of abstracts and index terms (with a controlled vocabulary and a controlled syntax) in order to supply users of the English, French, German or Spanish language with abstracts in their native language from inputs in one of the other languages
C7820; C7240
information analysis;language translation
TITUS II;international cooperative documentation system;Textile Information Treatment Users Service;automatic translation;abstracts;index terms;English;French;German;Spanish

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-LS-171	ISBN 92-835-0568-9	UNCLASSIFIED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly sur Seine, France

6. Title BENEFITS OF COMPUTER ASSISTED TRANSLATION TO INFORMATION MANAGERS AND END-USERS

7. Presented at

8. Author(s)/Editor(s)	9. Date
Various	June 1990

10. Author's/Editor's Address	11. Pages
Various	154

12. Distribution Statement This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

13. Keywords/Descriptors

Machine translation
Information centers
Information scientists
Utilization

14. Abstract

The aim of this Lecture Series is to show how computer assisted translation (CAT) can be of benefit not only to information managers but also to end-users. Existing systems will be described as well as the nature of the texts to be processed, the technical and human problems related to the use of such systems and the needs of end-users (quality level of translations, information acquisition in the mother tongue...). Examples of on-going applications and systems under development will also be presented. These examples will highlight the benefits documentation centres will derive from CAT and suggest solutions of interest to the end-user.

This Lecture Series, sponsored by the Technical Information Panel of AGARD, has been implemented by the Consultant and Exchange Programme.

AGARD Lecture Series No.171
Advisory Group for Aerospace Research and Development, NATO
BENEFITS OF COMPUTER ASSISTED TRANSLATION TO INFORMATION MANAGERS AND END-USERS
Published June 1990
154 pages

The aim of this Lecture Series is to show how computer assisted translation (CAT) can be of benefit not only to information managers but also to end-users. Existing systems will be described as well as the nature of the texts to be processed, the technical and human problems related to the use of such systems and the needs of end-users (quality level of translations, information acquisition in the

P.T.O.

AGARD-LS-171

Machine translation
Information scientists
Information centers
Utilization

AGARD Lecture Series No.171
Advisory Group for Aerospace Research and Development, NATO
BENEFITS OF COMPUTER ASSISTED TRANSLATION TO INFORMATION MANAGERS AND END-USERS
Published June 1990
154 pages

The aim of this Lecture Series is to show how computer assisted translation (CAT) can be of benefit not only to information managers but also to end-users. Existing systems will be described as well as the nature of the texts to be processed, the technical and human problems related to the use of such systems and the needs of end-users (quality level of translations, information acquisition in the

P.T.O.

AGARD-LS-171

Machine translation
Information scientists
Information centers
Utilization

AGARD Lecture Series No.171
Advisory Group for Aerospace Research and Development, NATO
BENEFITS OF COMPUTER ASSISTED TRANSLATION TO INFORMATION MANAGERS AND END-USERS
Published June 1990
154 pages

The aim of this Lecture Series is to show how computer assisted translation (CAT) can be of benefit not only to information managers but also to end-users. Existing systems will be described as well as the nature of the texts to be processed, the technical and human problems related to the use of such systems and the needs of end-users (quality level of translations, information acquisition in the

P.T.O.

AGARD-LS-171

Machine translation
Information scientists
Information centers
Utilization

AGARD Lecture Series No.171
Advisory Group for Aerospace Research and Development, NATO
BENEFITS OF COMPUTER ASSISTED TRANSLATION TO INFORMATION MANAGERS AND END-USERS
Published June 1990
154 pages

The aim of this Lecture Series is to show how computer assisted translation (CAT) can be of benefit not only to information managers but also to end-users. Existing systems will be described as well as the nature of the texts to be processed, the technical and human problems related to the use of such systems and the needs of end-users (quality level of translations, information acquisition in the

P.T.O.

AGARD-LS-171

Machine translation
Information scientists
Information centers
Utilization

mother tongue…). Examples of on-going applications and systems under development will also be presented. These examples will highlight the benefits documentation centres will derive from CAT and suggest solutions of interest to the end-user.

This Lecture Series, sponsored by the Technical Information Panel of AGARD, has been implemented by the Consultant and Exchange Programme.

ISBN 92-835-0568-9

mother tongue…). Examples of on-going *applications and systems under development* will also be presented. These examples will highlight the benefits documentation centres will derive from CAT and suggest solutions of interest to the end-user.

This Lecture Series, sponsored by the Technical Information Panel of AGARD, has been implemented by the Consultant and Exchange Programme.

ISBN 92-835-0568-9

mother tongue…). Examples of on-going applications and systems under development will also be presented. These examples will highlight the benefits documentation centres will derive from CAT and suggest solutions of interest to the end-user.

This Lecture Series, sponsored by the *Technical Information* Panel of AGARD, has been implemented by the Consultant and Exchange Programme.

ISBN 92-835-0568-9

mother tongue…). Examples of *on-going applications and systems under development* will also be presented. These examples will highlight the benefits documentation centres will derive from CAT and suggest solutions of interest to the end-user.

This Lecture Series, sponsored by the Technical Information Panel of AGARD, has been implemented by the Consultant and Exchange Programme.

ISBN 92-835-0568-9